Experiments in Fluids (2006) 40: 514–522
DOI 10.1007/s00348-005-0091-8

RESEARCH ARTICLE

Carlos M. Garcia · P. Ryan Jackson
Marcelo H. Garcia

Confidence intervals in the determination of turbulence parameters

Received: 28 July 2005 / Revised: 17 October 2005 / Accepted: 23 November 2005 / Published online: 13 December 2005

Abstract This paper quantifies uncertainty in turbulence parameters using the moving block bootstrap technique (MBB). This technique provides a good approximation of confidence intervals of turbulence parameters such as time and length scales of the turbulent flow, convective velocities and dissipation of turbulent kinetic energy which cannot be estimated based on the available standard methodology because of their computational complexity and reliance on correlation structure in the signal. The validation consisted of a comparison between uncertainty estimates of various parameters from MBB and those from analytical and experimental data.

1 Introduction

Present measurement technology applied to fluid dynamics studies allows researchers to capture water velocity signals with a relatively high sampling frequency. From these signals, the computation of parameters which describe the flow turbulence can be performed with good accuracy. The common turbulence parameters required in studies related to environmental and hydraulic engineering applications are: turbulent kinetic energy (TKE), Reynolds stresses, convective velocity (U_c), rate of dissipation of the TKE (ε), and turbulence scales in the flow such as the scales of the energy containing eddies (L) and the Kolmogorov eddies (η). These parameters can be used to understand the physics of the flow and to validate theories or numerical models.

It is quite common for experimentalists to present their results without discussing or estimating the errors involved. Error in a measurement is defined as the difference between its true value and its measured value and the magnitude of this error is difficult to quantify because the true value is unknown.

The error involved in the experimental determination of a turbulence parameter presents several components: errors due to the experimental setup (i.e., misallocation of the instrument); errors due to the physical constraint of the measurement technique (i.e., size of the sampling volume and sampling frequency of the instrument); statistical errors due to sampling a random signal, and errors due to the methodology used to compute the parameters (e.g. scaling relations provide only orders of magnitude of the parameter). This paper focuses on quantifying the statistical errors of the turbulence parameters which generate scatter in repeated measurement of parameters when the same experimental setup, instrument and methodology are used. The estimation of this error component will define an estimate of the lower bound of the total error and will provide a good estimate of the total error when the other error components are minimized and relatively small compared to the sampling error. Uncertainty analysis provides the experimentalist with a rational way of evaluating the significance of the statistical error component on repeated trials which can be used when comparing the reported data with results from other laboratory or numerical experiments. The interval around the measured value within which the true value is believed to lie is called the confidence interval. Statistical theory allows one to make the following inference (Efron and Tibshirani 1993): the true value of the parameter θ lies in the interval $XX \leq \theta \leq YY$ with a $100(1-2\alpha)\%$ confidence which basically consists of defining the confidence interval for θ. The coverage property of this interval implies that

C. M. Garcia (✉) · P. R. Jackson · M. H. Garcia
Ven T. Chow Hydrosystems Laboratory,
Department of Civil and Environmental Engineering,
University of Illinois at Urbana-Champaign,
205 North Mathews Ave, 61801 Urbana, IL, USA
E-mail: cgarcia2@uiuc.edu
Tel.: +1-217-3336183
Fax: +1-217-3330687

100(1−2α)% of the time, the interval $XX \leq \theta \leq YY$ will constrain the true value of θ. XX and YY are the 100 αth and 100(1−α)th percentiles of the θ distribution, respectively.

An approximation for this interval can be computed using a standard technique (Moffat 1988) assuming a normal probability distribution of the parameter which generates a symmetric interval around the mean value of the parameters. It is sometimes a good approximation, but it cannot be assumed a priori for all the turbulence parameters. The standard technique also requires an estimate of the standard error of the parameter (the standard deviation of the sampling distribution of that parameter). Equations are available in the literature to estimate the standard error for the mean, variances, Reynolds stresses, correlation coefficient, and third and fourth order moments, for *uncorrelated* samples presenting normal or arbitrary probability distribution (Benedict and Gould 1996). Bendat and Piersol (2000) presented a general equation for computing the standard error of the mean and the variance of correlated samples. However, no equations are available to compute the standard error of parameters which present more complicated expressions than sample mean and sample variance such as time and length scales, convective velocities, dissipation rate of TKE and Kolmogorov length scale. Such parameters require one to resolve and preserve the correlation structure of the turbulent velocity field.

Another method available in the literature, known as bootstrap and commonly used in econometrics, is used in the present work to approximate the confidence intervals for all of the turbulence parameters. The bootstrap method, introduced by Efron (1979), is based on resampling algorithms of the recorded signal. Bootstrap method overcomes most of all the cited standard technique deficiencies; it requires no theoretical calculations and it is available no matter how complicated the parameter may be (Efron and Tibshirani 1993). As it was introduced initially, the bootstrap method was restricted to deal with uncorrelated data. However, the moving block bootstrap technique (hereafter MBB), proposed by Kunsch (1989), is an option used during resampling to avoid destroying the real correlation in the signal which is being characterized. Such a technique can be applied to quantify the statistical uncertainty in turbulence parameters which require preservation of correlation structure in the measured signal. However, MBB has not been validated to approximate confidence intervals for turbulence parameters.

Section 2 of this paper provides background on the above cited techniques to approximate the confidence intervals (standard technique and bootstrap, focusing on MBB). Then, in Sect. 3, the MBB method for the definition of the confidence interval of the turbulence parameters is validated against analytical and experimental data for parameters such as sample mean and sample variance, turbulent kinetic energy, energy containing eddies scale, convective velocities, dissipation rate of TKE and Kolmogorov length scale. The first stage consists of the application of the MBB technique to a set of synthetic water velocity signals representing turbulence process for a wide range of flow conditions. The methodology used to build these synthetic signals is detailed in Garcia et al. (2005). The standard errors of the estimate of sample mean and variance of the generated signals computed using MBB were compared with the results available in the literature (Bendat and Piersol 2000). A set of laboratory experiments was performed in order to validate the use of moving block bootstrap to estimate the standard error of turbulence parameters which have more complicated forms than sample mean and variance. Basically, it consists of a set of 80 repeatable measurements of the three dimensional water velocity components recorded at the same location. All the turbulence parameters were computed for each of the 80 recorded signals, then the error variance and the probability distribution of each parameter were computed using the set of 80 values of each parameter. These values were compared with the values obtained by applying the MBB to one of the recorded signals.

2 Background

2.1 Standard technique to approximate confidence intervals

An approximation for the confidence interval of a defined turbulence parameter can be computed using the standard technique with coverage probability equal to (1−2α) or confidence level 100(1−2α)% which is usually defined as: $\theta \pm z^{(1\text{-}\alpha)}\ \sigma_e$, where σ_e is an estimate of the standard error for θ and $z^{(1\text{-}\alpha)}$ indicates the 100(1−2α) percentile of a normal distribution with zero mean and variance equal to unity as given in a standard normal table (Efron and Tibshirani 1993). The assumption of a normal distribution for the parameter which generates a symmetric interval around the mean value of the parameter is sometimes a good approximation, but it should be validated in each case.

Regarding the computation of the standard error σ_e, Benedict and Gould (1996) proposed methods for calculating the error variance σ_e^2 due to sampling of random process for different turbulence parameters such as the mean, variances, Reynolds stresses, correlation coefficient, and third and fourth order moments. In particular, formulae were proposed for σ_e^2 which are based on normal distribution assumptions and on any general distribution shape. However, these formulae only can be used on *uncorrelated* samples which require independence between samples. Practically speaking, the sampling frequency must be less than twice the inverse of the integral time scale of the quantity being measured. The proposed estimates of the variance errors of the sample mean and variance for *uncorrelated* samples valid for any parameter distribution are:

$$\sigma^2_{\text{mean}} = \frac{\sigma^2_u}{N} \tag{1}$$

$$\sigma^2_{\text{var}} = \frac{\mu_4 - \left(\sigma^2_u\right)^2}{N}, \tag{2}$$

where σ^2_u and μ_4 are the sample variance and fourth order moment of the variable u, respectively, and N is the total number of samples. Using the normal probability distribution assumption ($\mu_4 = 3\sigma^2_u$) the estimates of the error variance for the mean remain the same, but the error variance for the variance is:

$$\sigma^2_{\text{var}} = \frac{2\left(\sigma^2_u\right)^2}{N}. \tag{3}$$

Bendat and Piersol (2000) presented a general equation for computing the error variance of the mean and the variance of a *correlated* sample:

$$\sigma^2_{\text{mean}} = \frac{\sigma^2_u}{N} + \frac{2}{N^2}\sum_{k=1}^{N-1}(N-k) \times C_{uu}(k\Delta t) \tag{4}$$

$$\sigma^2_{\text{var}} = \frac{2}{N}\left[\left(\sigma^2_u\right)^2\right] + \frac{4}{N^2}\sum_{k=1}^{N-1}(N-k) \times C^2_{uu}(k\Delta t), \tag{5}$$

where $C_{uu}(k\Delta t)$ is the autocovariance function of the variable u

$$C_{uu}(k\Delta t) = R_{uu}(k\Delta t) - \mu^2_u, \tag{6}$$

μ_u is the sample mean of the variable u and $R_{uu}(k\Delta t)$ is the autocorrelation function of the variable u

$$R_{uu}(k\Delta t) = E[u(t) \times u(t + k\Delta t)]. \tag{7}$$

Two cases simulating different sampling configurations of a random process can be analyzed using these equations. Case 1: $T_x \ll \Delta t \ll T_m$ and Case 2: $\Delta t \ll T_x \ll T_m$ where $\Delta t = 1/\text{sampling rate}$, T_x is the integral time scale of the signal, and T_m is the total sampling time. For case 1, the samples are uncorrelated and $C_{uu}(k\Delta t) = 0$ for $k \geq 1$ implying that Eqs. 4 and 5 reduce to Eqs. 1 and 3 for those cases. For case 2 (conditions commonly found in water velocity signals recorded for turbulence characterization purposes) Eqs. 4 and 5 reduce to:

$$\sigma^2_{\text{mean}} = \frac{\sigma^2_u}{N_{\text{mean eff}}} \tag{8}$$

$$\sigma^2_{\text{var}} = \frac{2}{N_{\text{var eff}}}\left[\left(\sigma^2_u\right)^2\right], \tag{9}$$

where $N_{\text{mean eff}} = T_m/2T_x$ and $N_{\text{var eff}} = T_m/T_x$ are the number of effective samples for the sample mean and variance computation, respectively. Benedict and Gould (1996) stated that if the sampling rate is too high to ensure independence of samples, Eq. 1 through Eq. 3 should be used and N should be adjusted to reflect the number of integral time scales in the sampling time, not the actual number of samples. Otherwise, calculated confidence intervals based on N would be narrower than they actually are. However, they did not specify if the number of effective samples is different for each analyzed parameter. As seen above, the number of effective samples for the mean and variance is different ($N_{\text{mean eff}} \neq N_{\text{var eff}}$).

No equations like 4 and 5 are available to compute the error variance of parameters which present more complicated expressions than sample mean and sample variance such as time and length scales, convective velocities, dissipation rate of TKE and Kolmogorov length scale. Estimates of the error variance of the TKE were computed by Gould et al. (1997) using a root mean square combination of the error variance of each velocity component (u, v, w). This approach is valid when the variances are uncorrelated, repeated observations of each measurement, if made, display Gaussian distribution and the uncertainty in each measurement is initially expressed at the same odds (Moffat 1988).

2.2 Bootstrap technique to approximate confidence intervals

2.2.1 General description

The bootstrap method, introduced by Efron (1979), is a computer based method to estimate the standard errors and the probability distribution of the statistical parameters. Implementation of the bootstrap method consists of drawing randomly with replacement a number B of independent bootstrap series from the original data set with N data values each. For each bootstrap series, a bootstrap replication of each turbulence parameter is computed. Then, this information is used to infer the probability distribution of each parameter which allows the definition of confidence intervals without any assumption about the probability distribution of the parameter. Like any statistical method, bootstrap estimates are random variables with an inherent associated error.

The bootstrap method introduced by Efron (1979) was restricted to deal with uncorrelated data. Lately, several efforts have been made by statisticians to extend the bootstrap use to correlated data (Carlstein 1986; Kunsch 1989; Hall et al. 1995; Politis and White 2004). The MBB proposed by Kunsch (1989), is an option to use in these cases (i.e., turbulence) which avoids destroying the real correlation in the signal that one is trying to characterize.

2.2.2 The MBB technique

For a signal with N samples, the MBB technique proposed by Kunsch (1989) consists of choosing a block of a defined length b and considering all the possible contiguous blocks of this length ($N-b+1$ total blocks). One then randomly samples N/b blocks with replacement from the total number of blocks and pastes them together to form the bootstrap time series (see Fig. 1, block length equal to three samples). Enough blocks are sampled to obtain series of roughly the same length as the original series. This procedure is repeated B times to get B bootstrap signals. For each bootstrap signal the bootstrap replication of the desired statistics is computed and the probability distribution of a given parameter is estimated. Using this information, not only can the error variance of each parameter be computed, but so can the percentiles required to define the confidence intervals. Efron and Tibshirani (1993) stated that very seldom are more than $B = 200$ replications needed for estimating an error variance. However, much higher values of B (around 1,000) are required for bootstrap confidence intervals.

A key parameter in the MBB technique is the optimum length of the block. Politis and White (2004) proposed a methodology to choose the expected optimum block length (b_{opt}) by minimizing the mean square error (MSE) of the variance of the sample mean parameter. Their definition of the b_{opt} is based on the observed correlation structure $R(k\Delta t)$ of the recorded signal. The first step in the process of defining the optimum length of the moving block consists of identifying the smallest integer m after which the autocorrelation function appears negligible ($R(k\Delta t) \approx 0$ for $k > m$). Then, with $M = 2m$, b_{opt} is computed as:

$$b_{\mathrm{opt}} = \left[\left(\frac{2G^2}{D} \right)^{1/3} N^{1/3} \right], \tag{10}$$

where $D = 4/3\ g^2(0)$ and

$$g(0) = \sum_{k=-M}^{M} \lambda(k/M) \times R(k\Delta t) \tag{11}$$

$$G = \sum_{k=-M}^{M} \lambda(k/M) \times |k| \times R(k\Delta t). \tag{12}$$

In Eqs. 10 and 11, $R(k\Delta t)$ is defined in Eq. 7 and $\lambda(k/M) = 1$ if $0 \le |(k/M)| \le 0.5$, $\lambda(k/M) = 2(1 - |k/M|)$ if $0.5 \le |(k/M)| \le 1$ and $\lambda(k/M) = 0$ otherwise.

2.2.3 *Optimum length of the moving block for resampling turbulence signals using MBB*

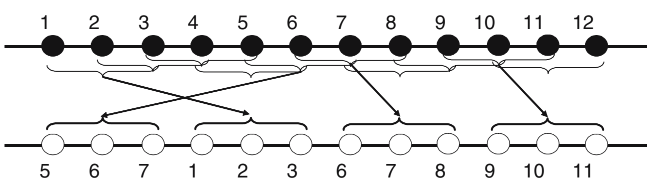

Fig. 1 A schematic diagram of the MBB for time series using a block length equal to three samples. The *black series* represents the original time series and the *white circles* denote one of the bootstrap realizations using MBB. *Numbers indicate* the sample order in the original signal (Adapted from Efron and Tibshirani 1993)

The methodology proposed by Politis and White (2004) is used herein to define the optimum length of the moving block for bootstrapping turbulence signals. An exponential model for the autocorrelation function of a velocity time signal in a turbulent flow is adopted herein as:

$$R_{xx}(k\Delta t) = \mathrm{e}^{-\left(\frac{k\Delta t}{T_x}\right)}. \tag{13}$$

If the water velocity signal, which generates the exponential autocorrelation function represented in Eq. 13, is sampled during $T_m = 2$ min with a sampling frequency, $f = 200$Hz ($N = 24{,}000$), the parameter m after which the autocorrelation function appears negligible is adopted here as $m = T_x/\Delta t = 20$ and $M = 2m = 40$. Thus, using Eqs. 10, 11, and 12, the optimum block length for bootstrapping this turbulent process is $b_{\mathrm{opt}} = 0.82$ s.

Thirty-seven scenarios were analyzed including different sampling configurations (25 Hz $\le f \le$ 200 Hz; 0.5 min $\le T_m \le$ 4 min) and exponential autocorrelation functions representing a wide set of flow conditions (0.04 s $\le T_x \le$ 10 s). The computed values of the optimum block length for each condition are included in Fig. 2.

Figure 2 also shows a fit of a power relation between the optimum block length, b_{opt} and the sampling time T_m and the integral time scale of the velocity signal T_x as:

$$b_{\mathrm{opt}} = 0.788 \times T_x^{2/3} T_m^{1/3}. \tag{14}$$

Since all turbulence parameters are computed from the same velocity signal, the optimum block length for the sample mean, given either by Eq. 10 through Eq. 12 or by Eq. 14, is used in the application of the MBB technique for all the turbulence parameters.

3 Validation of moving block bootstrap

3.1 Using analytical data

A set of 32 synthetic turbulent signals, generated following the methodology of Garcia et al. (2005), was used herein to validate the complete MBB technique (including the definition of the optimum block length). Synthetic data were used to simulate a broad set of flow conditions which could be very difficult to attain in laboratory experiments. The synthetic data set included 2 min long one dimensional velocity signals for a wide range of flow conditions (0.1 m/s $\le U_c \le$ 1 m/s, 0.1 m $\le L \le$ 1 m, and 0.0001 m $\le \eta \le$ 0.0005 m). First, the MBB technique was applied to each time signal using $B = 1{,}000$ replications. Thus, the error variance of the sample mean and variance were computed for each signal and the values were contrasted in Figs. 3 and 4 against the values obtained by using the analytical equations 8 and 9, respectively.

A very good agreement was observed between MBB and the analytical equations which validate the use of

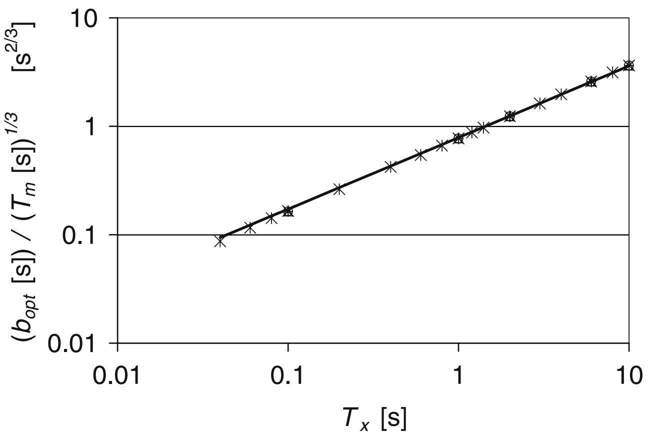

Fig. 2 The optimum block length, b_{opt} as a function of the sampling time T_m and the integral time scale of the velocity signal T_x

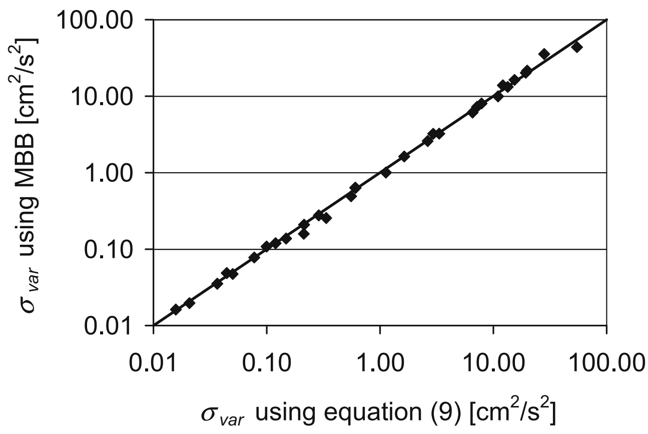

Fig. 4 Comparison between standard errors of the sample variance computed for each signal using MBB and Eq. 9. The *continuous line* represents the perfect agreement between the two options

the MBB to compute the error variance for sample mean and variance. For flow conditions where the ratio $T_m/T_x \geq 20$ ($T_m < 20T_x$ do not provide a good estimate of the sample mean according to Garcia et al. 2004), the maximum relative error of the standard error of the sample mean is 9% with an average value of 2%. For the sample variance, the maximum relative error was 26% with an average value of 8%.

Using the synthetic signals, all the turbulence parameters (U_c, TKE, L, ε, η, and T_x) were computed according to the methodology described in the Appendix. The error variances or standard errors for all of these parameters were computed using MBB but no analytical equations were available to compare them with. The MBB technique is validated for the computation of the standard error of the turbulence parameters using experimental data in the next section and an analysis of the behavior of the standard error of these parameters as the flow and sampling configuration change is presented.

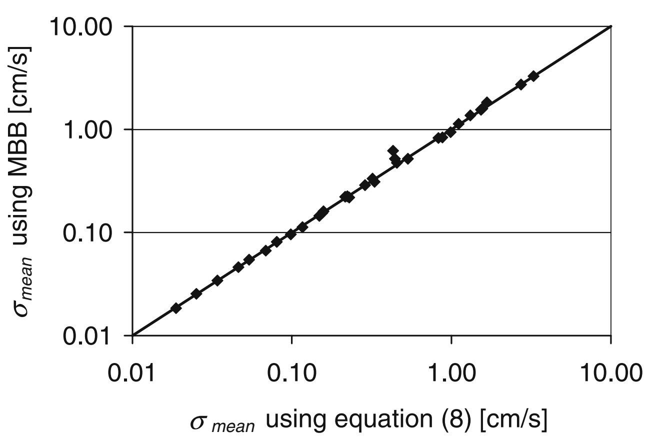

Fig. 3 Comparison between standard errors of the sample mean computed for each synthetic signal using MBB and Eq. 8. The continuous line represents perfect agreement between the two options

3.2 Using experimental data

Validation of the MBB method for turbulence parameters which present more complicated analytical forms than the sample mean and variance (for instance TKE, L, ε, η and T_x) was performed using experimental data. As it was mentioned before, MBB predicts statistical errors that should be obtained on repeated trials with the same experimental set-up, same experimental conditions, same instrumentation and same methodology. Those conditions are reproduced here. The validation roughly consists of performing a repeatable set of measurements of water velocity signals a certain number of times, computing the turbulence parameters for each one of these signals following the methodology included in the Appendix, and from them finally estimating the probability distribution of each parameter. The error variance values and the confidence interval are then compared with the values obtained using MBB.

3.2.1 Facility and methods

The experimental facility chosen for these experiments was selected for its ability to maintain a stationary and fully developed turbulent flow for an extended period of time with no significant changes in mean fluid and flow properties in order to minimize the deterministic errors present in the recorded signal. Preliminary experiments were completed in a small tilting flume, but the recirculation of the water through the facility pumps increased the mean temperature of the water by about 4° over the course of the experiments (4 h). Variations in water temperature during the experiments leads to biases in the results of the study, so another facility with a minimal influence on temperature during recirculation and housed in a controlled environment was used for the experiments.

The experimental facility used in these experiments was an Odell–Kovasnay type recirculation flume (Odell and Kovasnay 1971) built by Engineering Laboratory Design (ELD) in Lake City, Minnesota (see Fig. 5). This closed loop water channel is primarily used for stratified flows experiments, but the very small thermal conductance from the disk pump to the test fluid and the insulating properties of the flume provided excellent flow conditions for the present experiments. To ensure that the test fluid was at equilibrium with the room temperature prior to the experiments, the facility was filled to a total depth of 54 cm with tap water 3 days prior to the experiment and allowed to sit with no disturbances.

An acoustic Doppler velocimeter (ADV) was used to record time series of the three velocity components at one point in the flow field. The ADV used to gather the data was a SONTEK MicroADV sampling at 25 Hz with a velocity range of 10 cm/s. The measurement volume of the ADV (cylinder of water with a diameter of 4.5 mm and a height of 5.6 mm) was located at 17.3 cm from the downstream end of the test section along the center line, 7.5 cm from the channel walls, and 43.5 cm above the bottom. The ADV was mounted on a Velmex Unislide set on an instrument cart which allowed precise horizontal and vertical positioning of the instrument as well as a mount free from vibrations. Seeding material provided by the ADV manufacturer was used to improve the acoustic backscattering in the flow. Garcia et al. (2005) analyzed the uncertainty on the determination of the turbulence parameters due to physical constraints of the acoustic Doppler Velocimetry technique and they found that for certain flow conditions and instrument configurations (defined by a dimensionless number based on the sampling frequency, energy containing eddy length scale and convective velocities in the flow), the physical constraint component of the uncertainty could be considered negligible. The present experiments satisfy their requirements.

Preliminary data was gathered for 1 h to insure no changes in temperature and that the system reached a steady state. This preliminary data also provided an estimate of the integral time scale in the velocity signals of about 1 s. Based on this value, a 2 min sampling period was selected a priori as the optimum sampling length for the analyzed experimental conditions. This length of the record ensures that more than 100 of the largest turbulence structures were averaged during the sampling time. Once the system was in steady state, a total of 80 2-min data sets were gathered over 5 h. The two parameters commonly used to quantify the quality of the water velocity signals recorded using ADV are the correlation score and the signal to noise ratio. The mean (97% and 22.8 dB) and minimum (96.7% and 20.0 dB) values of these parameters observed for the set of 80 signals are much higher than the minimum values usually required by the manufacturer to record signals for turbulence characterization purposes (70% and 15dB, respectively). The air temperature in the lab increased only 0.5°C and the water temperature gained a mere 0.1°C during the 5 h it took to complete the experiment.

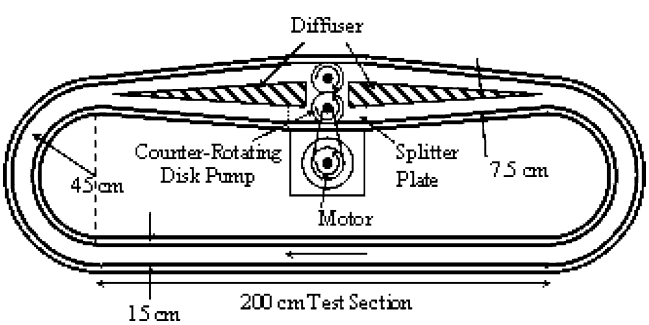

Fig. 5 Odell–Kovasnay recirculation facility used in the experimental validation of the MBB

3.2.2 *Results*

The mean velocities and all the turbulence parameters (TKE, L, T_x, U_c, ε and η) were computed for each recorded time signal using the methodology described in the Appendix. Noise energy levels (NEL) of the recorded signals were detected and corrections to all the turbulence parameters due to this noise were performed according to the technique detailed by Garcia et al. (2005). The analyzed flow conditions (averaging the 80 signals) at the point where the measurements were performed were: mean streamwise (longitudinal) flow velocity of 7.45 cm/s, transversal and vertical mean velocities smaller than 0.5 cm/s, integral time scale in the longitudinal direction T_x equal to 0.96 s, integral length scale in the longitudinal direction $L_x = 7.32$ cm, TKE = 0.83 cm^2/s^2, dissipation rate of TKE $\varepsilon = 0.066$ cm^2/s^3 and Kolmogorov length scale $\eta = 0.063$ cm.

The error variances of each turbulence parameter (as well as the noise energy level) were computed using the set of 80 computed values of each parameter. These values of error variance were compared with the equivalent values obtained for all the same parameters applying the MBB ($B = 1{,}000$ replications) to one of the recorded signals using the optimum moving block length defined by two available options: (a) using Eqs. 10, 11, and 12 with the observed autocorrelation function and (b) using Eq. 14 based on the exponential model of the autocorrelation function and the observed value of T_x.

The observed autocorrelation function R_{xx} (see Fig. 6) is required to estimate the optimum moving block length b_{opt} using Eqs. 10, 11, and 12. Assuming that the turbulence is the main random process in the recorded signal and that the exponential model of R_{xx} adequately represents the random process, Eq. 14 can be used (using the mean observed value of $T_x = 0.96$ s). Figure 6 shows a good agreement between the observed values of the longitudinal velocity autocorrelation function (averaged over the 80 signals) and the exponential model for the recorded flow conditions. The values of b_{opt} computed using the observed and the exponential model autocorrelation function did not

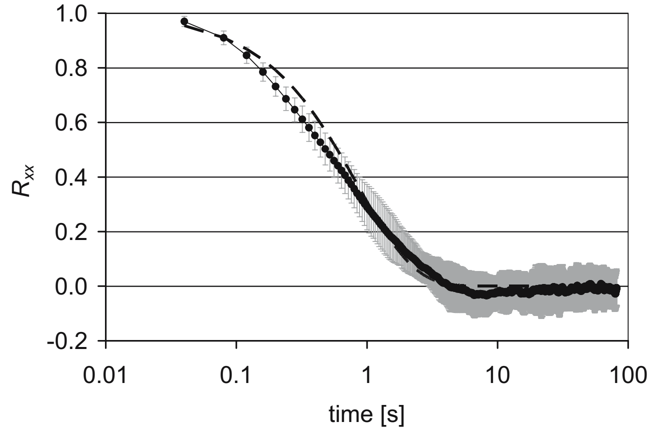

Fig. 6 Comparison between observed and simulated (exponential model) autocorrelation functions for the analyzed flow conditions. The *continuous line* represents the observed values (averaged over the 80 signals) and the *dashed line* shows the exponential model. *Error bars* indicate one standard deviation away from the average value based on the 80 recorded signals

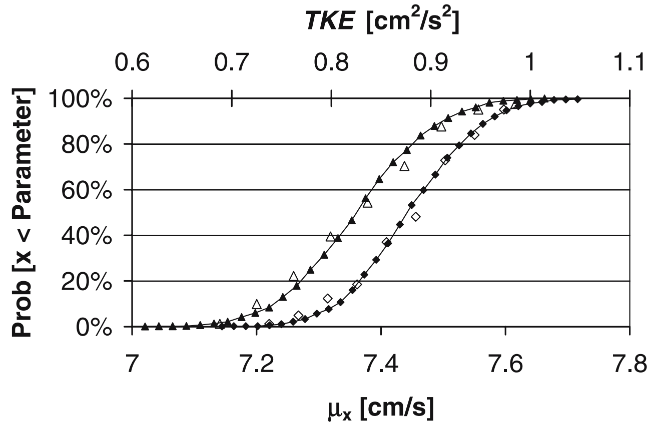

Fig. 7 Comparison between probability distribution of the sample mean (*filled diamond*, *diamond*) and TKE (*filled triangle*, *triangle*) estimated from the experimental data (*open symbols*) and by using MBB with b_{opt} from Eq. 14 (*dark symbols*)

significantly differ and they were 3.91 and 3.78 s, respectively.

Table 1 compares the coefficients of variation of each turbulence parameter computed from the set of repeatable measurements (experimental data) and by using MBB (block length computed using Eq. 14). The coefficient of variation is defined here as the ratio between the standard error computed using MBB and the mean value of the parameter. Very good agreement is observed between values obtained from the experimental data and by using MBB with b_{opt} estimated with Eq. 14. For the sample mean, variance and convective velocity the relative differences are smaller than 6%. The other turbulence parameters show differences lower than 20% (only the noise energy level NEL_x is around 24%). It is important to note that the Kolmogorov length scale η

Table 1 Coefficient of variation of each turbulence parameter obtained from the set of repeatable measurements (experimental data) and by using MBB

Parameter	Observed (80 repeatable measurements) (%)	MBB (b_{opt} from Eq. 14) (%)
Mean (μ_x)	1.33	1.25
Variance (σ_x^2)	13.05	12.69
TKE	8.96	7.55
Convective velocity (U_{cx})	1.35	1.31
Integral time scale (T_x)	42.70	34.04
Dissipation rate of TKE (ε)	20.32	23.18
Kolmogorov length scale (η)	4.99	5.94
Noise energy level (NEL_x)	12.40	9.35

The subscript x denotes that the parameter corresponds to the streamwise (longitudinal) flow direction

was estimated here using a scaling relation. Therefore, the η estimate only provides an order of magnitude estimate of the size of the dissipating eddies. Thus, the variation observed for this parameter in the 1,000 MBB replications (between 0.05 and 0.08 cm with a coefficient of variance of about 6%) are all the same order of magnitude, hence they are the same result. Thus, the uncertainty in the selected methodology to determine η is much higher than statistical errors (reproduced by MBB).

In addition to the variance error of the turbulence parameter, the MBB technique also provides information about the probability distribution of each parameter which is estimated from the 1,000 MBB replications (at least 1,000 replications are required for the description of the probability distribution of each parameter according to Efron and Tibshirani 1993). Figures 7 and 8 show the accuracy of MBB (1,000 replications) at representing the observed probability distribution from the experimental data. Corrections due to presence of bias between the observed and the MBB probability distributions were performed before such comparison. The cited bias is due to the differences between the average values computed from the 80 repeatable measurements and from the 1,000 bootstrap replications. The bias values for all the parameters included in Figs. 7 and 8 were $\leq 7\%$ of the parameter average value computed from the 80 velocity signals. A good performance of the MBB (1,000 replications) using b_{opt} from Eq. 14 is observed in these figures reproducing the main features of the observed distribution obtained from the experimental data.

Using the probability distributions estimated using MBB, the confidence intervals can be defined for different significance level. The 5 and 95% percentiles of a parameter estimated using the MBB technique based on 1,000 replications are the 50th and 950th ordered value of these replications.

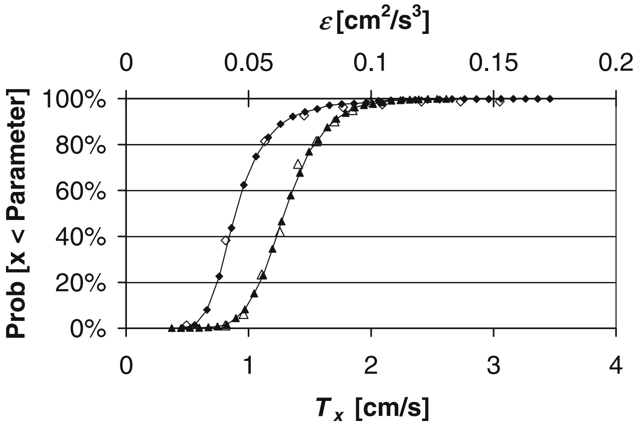

Fig. 8 Comparison between probability distribution of the integral time scale, T_x (*filled diamond*, *diamond*) and the dissipation rate of TKE (*filled triangle*, *triangle*) estimated from the experimental data (*open symbols*) and by using MBB with b_{opt} from Eq. 14 (*dark symbols*)

The estimated probability distribution of each parameter can be analyzed to test their normality. Normal distribution of the parameter is commonly used as a valid assumption for the standard technique to approximate the confidence intervals. Tests for goodness of fit to a normal distribution were performed for each estimated probability distribution. Two different tests were used herein: The Lilliefors test (Lillierfors 1967) and the Jarque–Bera test (Bera and Jarque 1981). Both tests evaluate the hypothesis that a defined parameter has a normal distribution with an unspecified mean and variance, against the alternative that this parameter does not have a normal distribution. According to these tests, the hypothesis of normal distribution could not be rejected (at a significance level of 5%) for mean velocity (μ_x), TKE, convective velocity (U_c), and noise energy level (NEL_x). On the other hand, the hypothesis of normal distribution was rejected according to these tests at the same significance level for the variance, integral time scale T_x, dissipation rate of TKE ε and Kolmogorov length scale (η). These parameters showed positive skew (also called skewed to the right) in their distribution which could be generated for the methodology used to estimate them (see the Appendix).

4 Summary and conclusions

The MBB technique, commonly used in econometrics, has been validated to provide a good approximation of the confidence intervals of turbulence parameters. This paper uses the MBB technique to quantify the statistical errors of the turbulence parameters which generate scatter in repeated measurement of the parameters when the same experimental setup, instrument and methodology are used.

The main contribution of this paper is related to the determination of uncertainties of turbulence parameters such as time and length scales, convective velocities and dissipation of TKE which could not be estimated based on the available standard methodology because of their computational complexity and reliance on correlation structure in the signal. A key parameter in the MBB technique is the optimum length of the moving block. An estimate of the block length was derived based on an exponential model for the autocorrelation function of a one dimensional velocity signal representing a turbulent process. This equation is not valid for water velocity signals which present random periodic processes superimposed on the turbulence process (i.e., tidal, internal waves, etc.). In those cases, the observed autocorrelation function should be used with the Politis and White (2004) technique to estimate the optimum moving block length.

The MBB method for the approximation of confidence intervals of the turbulence parameters was validated against analytical and experimental data. A good agreement was observed between MBB and analytical equations available in the literature to compute the error variance for sample mean and variance of water velocity signals representing a wide range of flow conditions. A set of laboratory experiments, performed in a facility housed in a controlled environment, validated the use of moving block bootstrap to estimate the error variance of turbulence parameters which have more complicated forms than sample mean and variance and rely on correlation structure. Good agreement is observed between the uncertainty of all analyzed turbulence parameters obtained from the experimental data and by using the MBB technique.

In addition to the agreement of the estimated values of the error variance, the MBB is observed to reproduce the main features of the probability distribution obtained from experimental data. Using the probability distributions estimated using MBB, the confidence intervals were defined for different significance levels without any assumption about the probability distribution of the parameter. According to two normality tests performed on the generated probability distribution, the hypothesis of normal distribution is rejected at the 5% significance level for the following parameters: variance, time scale of the energy containing eddies, rate of dissipation of TKE and Kolmogorov length scale. The presence of different probability distributions for each turbulence parameter gives more support to the use of MBB to approximate the confidence interval because this method is not based on the normality assumption.

Acknowledgments The research summarized in this paper resulted for work supported by the U.S. Geological Survey (Office of Surface Water), the Metropolitan Water Reclamation District of Greater Chicago and the Office of Naval Research. All this support is gratefully acknowledged. Comments from Dr. Arthur Schmidt and from two anonymous reviewers provided valuable insights.

5 Appendix: methodology used to compute the turbulence parameters

The TKE is computed by definition as TKE $= 1/2\,(u'^2 + v'^2 + w'^2)$ where u'^2, v'^2 and w'^2 indicate the variance (corrected by Doppler noise presence, Garcia et al. 2005) of the signal for each Cartesian velocity component. The convective velocity in the x Cartesian direction (with mean velocity U) is estimated following Heskestad (1965):

$$U_c^2 = U^2\left[1 + 2\frac{V^2}{U^2} + 2\frac{W^2}{U^2} + \frac{\overline{u'^2}}{U^2} + 2\frac{\overline{v'^2}}{U^2} + 2\frac{\overline{w'^2}}{U^2}\right] \quad (15)$$

where U, V and W are the mean flow velocities in the three Cartesian coordinates. It is valid for incompressible shear flows steady in the mean where the scale of non homogeneity is large compared to the Taylor microscale. Such flows include high Reynolds number, turbulent, free shear flow and wall-restricted shear flows away from the solid boundaries. The integral time scale T_x of a turbulent process is computed by integrating the autocorrelation function in the time domain to the first zero crossing according to Kresta and Wood (1993) and Wernersson and Trägårdh (2000). The integral length scale is computed using the convective velocity U_c value as $L_x = T_x U_c$. On the other hand, the order of magnitude of the smallest scale of turbulent motion (Kolmogorov length scale) is estimated using a scaling analysis as $\eta \propto (v^3/\varepsilon)^{1/4}$ (Pope 2000), where v is the kinematic viscosity and ε is the dissipation rate of turbulent kinetic energy. Finally, the rate of dissipation of TKE, ε is computed from $-5/3$ slope fitting of the observed power spectrum (generated from the complete signal) in the inertial range using the one dimensional power spectrum in the spatial domain: $E_{11}(k_1) = C_1\, k_1^{-5/3}\, \varepsilon^{2/3}$, where $C_1 = 0.49$ (Pope 2000). The wave number k_1 and the one-dimensional spectrum in the spatial domain E_{11} can be computed from information in the frequency domain using the Taylor hypothesis. The limits of the range where the non linear fit is performed are defined based on visual inspection of the power spectrum in order to use only the range where $-5/3$ slope is present. The true slope of the data in the selected range is computed as an indicator to verify the $-5/3$ assumption.

References

Bendat J, Piersol A (2000) Random data, 3rd edn. Wiley, New York

Benedict L, Gould R (1996) Towards better uncertainty estimates for turbulence statistics. Exp Fluids 22:129–136

Bera A, Jarque C (1981) Efficient tests for normality, homoscedasticity and serial independence regression residuals—Montecarlo evidence. Eco Lett 7:313–318

Carlstein E (1986) The use of subseries values for estimating the variance of a general statistic from a stationary sequence. Ann Stat 14:1171–1179

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Stat 7:1–26

Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman and Hall, New York

García C, Cantero M, Jackson P, García M (2004) Characterization of the flow turbulence using water velocity signals recorded by Acoustic Doppler Velocimeters. Civil engineering studies, Hydraulic engineering series 75. University of Illinois at Urbana-Champaign

García C, Cantero M, Niño Y, García M (2005) Turbulence measurements using acoustic Doppler velocimeter. J Hydraulic Eng 131 (in press)

Gould R, Benedict L, Raffoul C, Nejad A (1997). Statistical uncertainties of turbulent kinetic energy balances. ASME fluids engineering division summer meeting FEDSM'97. Vancouver, Canada

Hall P, Horowitz J, Jing B (1995) On blocking rules for the bootstrap with dependent data. Biometrika 82:561–574

Heskestad G (1965) A generalized Taylor hypothesis with application for high Reynolds number turbulent shear flows. J Appl Mech 32:735–739

Kresta S, Wood P (1993) The flow field produced by a pitched blade turbine. Characterization of the turbulence and estimation of the dissipation rate. Chem Eng Sci 48:1761–1774

Kunsch H (1989) The jackknife and the bootstrap for general stationary observations. Ann Stat 17:1217–1241

Lilliefors H (1967) On the Kolmogorov–Smirnov test for normality with mean and variance unknown. J Am Stat Assoc 62:399–402

Moffat R (1988) Describing the uncertainties in experimental results. Exp Therm Fluid Sci 1:3–17

Odell G, Kovasnay L (1971) A new type of water channel with density stratification. J Fluid Mech 50:535–543

Politis D, White H (2004) Automatic block-length selection for the dependant bootstrap. Econ Rev 23:53–70

Pope S (2000) Turbulent flows. United Kingdom, Cambridge

Wernersson E, Trägårdh C (2000) Measurements and analysis of high-intensity turbulent characteristics in a turbine-agitated tank. Exp Fluids 28:532–545