12th International Symposium on Turbulence and Shear Flow Phenomena (TSFP12)
Osaka, Japan (Online), July 19–22, 2022

ESTIMATING UNCERTAINTY OF LOW- AND HIGH-ORDER TURBULENCE STATISTICS IN WALL TURBULENCE

Saleh Rezaeiravesh, Donnatella Xavier & Ricardo Vinuesa
Department of Engineering Mechanics
KTH Royal Institute of Technology
Stockholm, Sweden
{salehr,dgxavier,rvinuesa}@mech.kth.se

Jie Yao
Department of Mechanical Engineering,
Texas Tech University
Lubbock, Texas, USA
jie.yao@ttu.edu

Fazle Hussain
Department of Mechanical Engineering,
Texas Tech University
Lubbock, Texas, USA
fazle.hussain@ttu.edu

Philipp Schlatter
Department of Engineering Mechanics
KTH Royal Institute of Technology
Stockholm, Sweden
pschlatt@mech.kth.se

ABSTRACT

A framework is introduced for accurate estimation of time-average uncertainties in various types of turbulence statistics. A thorough set of guidelines is provided to adjust the different hyperparameters for estimating uncertainty in sample mean estimators (SMEs). For high-order turbulence statistics, a novel approach is proposed which avoids any linearization and preserves all relevant temporal and spatial correlations and cross-covariances between SMEs. This approach is able to accurately estimate uncertainties in any arbitrary statistical moment. The usability of the approach is demonstrated by applying it to data from direct numerical simulation (DNS) of the turbulent flow over a periodic hill and through a straight circular pipe.

TIME-AVERAGE UNCERTAINTY

One of the main goals of any turbulent flow simulation is to compute various flow statistics after the statistically-stationary condition is achieved. Recent progress in hardware and software has made the scale-resolving simulations of turbulent flows more accessible. However, any such simulation can only be run for a finite time; therefore, the sample mean estimator (SME) $\hat{\mu} = \hat{\mathbb{E}}[u]$ for computing the time-average of a turbulent quantity u is uncertain. The SMEs are unbiased, and by ergodicity and the central limit theorem,

$$\hat{\mu} \sim \mathcal{N}\left(\mu, \sigma^2(\hat{\mu})\right) \tag{1}$$

where $\mathcal{N}$ is a Gaussian distribution with true mean $\mu = \mathbb{E}[u]$ and variance $\sigma^2(\hat{\mu})$. The latter is a direct measure of the uncertainty in $\hat{\mu}$ in a single realization of a turbulent flow where the associated μ cannot be realized. Since the time series of turbulence quantities are highly correlated, for estimating $\sigma^2(\hat{\mu})$ standard linear unbiased estimators typically used for independent and identically distributed (iid) samples are not accurate. The remedy can be found in the field of uncertainty quantification (UQ) where the tools for estimating $\sigma^2(\hat{\mu})$ for autocorrelated time samples can be divided into two main categories: 1. Batch-based methods, which are based on batching the time series and acting on the batch means, examples are non-overlapping/overlapping batch means (NOBM/OBM) methods and batch-means and batch-correlations method (BMBC) introduced by Russo & Luchini (2017). 2. Autocorrelation-function (ACF)-based estimators, which rely on evaluating the exact analytical expression for $\sigma^2(\hat{\mu})$:

$$\sigma^2(\hat{\mu}) = \frac{1}{n}\left[\gamma_0 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\gamma_k\right], \tag{2}$$

where $\gamma_k = \mathbb{E}[(u_i - \hat{\mu})(u_{i+k} - \hat{\mu})]$ is the autocovariance for the statistically-stationary sequence $\{u_i\}_{i=1}^n$ at $k = 0, 1, 2, \ldots$ time lags. As it can be inferred from Eq. (2), for a given number of samples two main factors which can give rise to the variance of an SME are the large variance of time series and the slow decay of the autocorrelation $\gamma(k)/\gamma(0)$. These two factors can be large either due to the physics of turbulence or numerical artifacts. An illustration of the latter can be found in Rezaeiravesh *et al.* (2022), where it was shown that the spatial location along the profile of a turbulence quantity where the sensitivity with respect to the numerical parameters is high, the time-average uncertainty can be large as well.

The main barrier to accurate estimation of $\sigma^2(\hat{\mu})$ via Eq. (2) is the oscillations in the sample-estimated autocovariances $\hat{\gamma}(k)$ at large lags. As first shown by Oliver *et al.* (2014), autoregressive models (ARM) are suitable for constructing smooth models for the ACF in turbulent flows (hereafter, ARM-based uncertainty estimator). An alternative to constructing an ARM is to fit an ad-hoc function to sample ACF, see e.g. Gscheidle *et al.* (2022). In our recent study, Xavier *et al.* (2022), the performance of various estimators with respect to associated hyperparameters was thoroughly investigated for wall turbulence. For batch-based variance estimators, the hyperparameter is the batch size, and we observed that for the NOBM, it could be determined from the lag-1 temporal correlation of the batch means. For the BMBC estimator, the lag-2 and lag-1 correlations of the batch means are both

needed for optimal estimation of batch size and hence accurate estimation of $\sigma^2(\hat{\mu})$. For modeled-ACF-based estimators, it is shown that the order of the ARM could be obtained from the turbulence integral time scale, without the need for any mathematical order selection criteria. By constraining the hyperparameters using integral quantities (such as eddy turnover times), guidelines were derived by Xavier *et al.* (2022) for accurate estimation of uncertainty in SMEs of primitive variables in turbulent flows.

HIGH-ORDER & NONLINEAR STATISTICS

The methods reviewed in the previous section can only be applied to estimate uncertainty in a single sample mean estimator. But, turbulence statistics include higher-order moments of primitive variables, nonlinear terms and different budget terms in the transport equations of Reynolds-stress components and turbulent kinetic energy. As outlined in this section, the present study proposes a UQ approach to accurately estimate time-average uncertainty in any turbulence statistics. Using combinations of arithmetic and differentiation operators, any of such statistical terms can be written in terms of the moments of the primitive variables. For instance, the second-order central moment $\langle u'v'\rangle$ can be written as $\langle u'v'\rangle = \langle uv\rangle - \langle u\rangle\langle v\rangle$. Hereafter, we denote a compound statistical term such as $\langle u'v'\rangle$ by T which is defined in terms of the first-order moments $\mathbf{M}$ which in this case are $\mathbf{M} = \{\langle u\rangle, \langle v\rangle, \langle uv\rangle\}$. In general, decomposition of T in terms of $\mathbf{M}$ should be in a way that the uncertainty in SMEs $\hat{\boldsymbol{\mu}} = \{\hat{\mu}_1, \hat{\mu}_2, \cdots, \hat{\mu}_p\}$ which correspond to the members of $\mathbf{M}$ can be estimated by any of the approaches of the previous section.

The problem to tackle is defined as estimating uncertainty in $T = f(\mathbf{M})$ where $f(\cdot)$ is a given functional, from estimated uncertainty in the SMEs of the members of $\mathbf{M}$. Clearly, this is a UQ forward (uncertainty propagation) problem, (Smith, 2013). For better understanding the features of the proposed approach, a brief review of the relevant studies for estimating uncertainty in compound statistics T is given first. The resampling methods exercised by Benedict & Gould (1996) do not consider the fact that turbulence time-series samples are autocorrelated; therefore, its resulting uncertainty estimations can be inaccurate. For the turbulent channel flow, Hoyas & Jiménez (2008) applied the non-overlapping batch method and Lee & Moser (2015) employed the ARM-based uncertainty estimator. However, these studies do not provide any detail on how uncertainties in high-order statistics are estimated.

Our proposed UQ approach has a number of distinguishing characteristics. First, all the arithmetic and differentiation operations corresponding to $f(\cdot)$ are applied in a mathematically and statistically consistent way while avoiding any approximation such as linearization. As a key feature, the correlations between different SMEs appearing in a compound term as well as the spatial correlations in the SMEs are accurately incorporated. Representing the covariance matrix associated with SMEs $\hat{\boldsymbol{\mu}} = \{\hat{\mu}_1, \hat{\mu}_2, \cdots, \hat{\mu}_p\}$ by $\boldsymbol{\Sigma}$, Eq. (1) can be generalized as,

$$\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3}$$

where the Gaussian distribution is p-variate. The diagonal elements of $\boldsymbol{\Sigma}$ are $\sigma^2(\hat{\mu}_i)$ for $i = 1, 2, \ldots, p$ and can be estimated by any of the methods in the previous section. The estimation of the off-diagonal elements which are the covariance between the SMEs has, to our knowledge, not been previously addressed in the literature. For these covariances the following exact expression can be derived,

$$\mathrm{cov}(\hat{\mu}_\mathrm{i}, \hat{\mu}_\mathrm{j}) = \frac{1}{\mathrm{n}}\left[\gamma_\mathrm{ij}(0) + \sum_{\mathrm{k}=1}^{\mathrm{n}-1}\left(1 - \frac{\mathrm{k}}{\mathrm{n}}\right)\left(\gamma_\mathrm{ij}(\mathrm{k}) + \gamma_\mathrm{ij}(-\mathrm{k})\right)\right] \tag{4}$$

where $\gamma_{ij}(k) := \mathrm{cov}(\mathrm{u_{i_l}}, \mathrm{u_{j_{l+k}}})$ is the cross-covariance between the time-series associated with two quantities u_i and u_j. Note that the only symmetry is $\gamma_{ij}(k) = \gamma_{ji}(-k)$. As a convenience, if $k > 0$ in $\gamma_{ij}(k)$, then k is a lead for u_i and a lag for u_j. Also, the peak of $\gamma_{ij}(k)$ does not necessarily happen at $k = 0$. Note that for $i = j$, Eq. (4) is reduced to Eq. (2). Similar to Eq. (2), the sample-estimated values of $\gamma_{ij}(\pm k)/\gamma_{ij}(0)$ can be oscillatory at high k, therefore, it is necessary to develop methods to construct models for these cross-covariances which are smooth functions of leads/lags. This can, in general, be more challenging compared to the modeling of $\gamma(k)$ in Eq. (2). An approach could be using an algebraic function as introduced in Gscheidle *et al.* (2022). A workaround, which can however be less accurate, is to cut off the summation in (4) at a k above which the sample-estimated $\gamma_{ij}(\pm k)$ are oscillatory.

To summarize, the steps of the algorithm for estimating uncertainty in $T = f(\mathbf{M})$ are the following:

1. Estimate uncertainty in $\hat{\boldsymbol{\mu}}$ corresponding to the SMEs of the members of $\mathbf{M}$ using any of the approaches in the previous section;
2. Estimate the off-diagonal elements of $\boldsymbol{\Sigma}$ from Eq. (4);
3. Using $\boldsymbol{\Sigma}$ and approximating $\boldsymbol{\mu}$ by $\hat{\boldsymbol{\mu}}$, directly draw N joint samples from Eq. (3) and estimate T (denoted by $\hat{T}$). Note that this is a pure Monte Carlo (MC) sampling method for the UQ forward problem;
4. For sufficiently large N, use sample estimator for mean and standard-deviation of $\hat{T}$. The latter is used to compute a confidence interval for the mean of $\hat{T}$.

The only approximation involved in this algorithm is in step 3, where $\boldsymbol{\mu}$ is replaced by $\hat{\boldsymbol{\mu}}$. If the number of time samples for estimating $\hat{\boldsymbol{\mu}}$ is sufficiently high, the potentially induced error in the estimated variance of $\hat{T}$ is negligible.

The above algorithm along with various time-average uncertainty estimators including those in the previous sections have been implemented in UQit, (Rezaeiravesh *et al.*, 2021) and will be published as open-source.

VALIDATION OF THE UQ ALGORITHM

The proposed UQ algorithm aims at estimating the uncertainty of the statistics in a single realization of any dynamical system when the statistically-stationary conditions hold. To validate the algorithm, its estimated uncertainties in various statistical terms can be compared to the corresponding ensemble uncertainties obtained from repeating a simulation (multi realizations). If a simulation is repeated N_e times (N_e is also called the size of ensemble) while keeping all simulations independent of each other, then the empirical ensemble mean and variance of a sample-mean estimation $\hat{\mu}$ are respectively obtained from the following two expressions:

$$\mu_e = \mathbb{E}[\hat{\mu}] \approx \frac{1}{N_e}\sum_{i=1}^{N_e}\hat{\mu}_i\,, \tag{5}$$

$$\sigma_e^2 = \mathbb{V}[\hat{\mu}] \approx \frac{1}{N_e}\sum_{i=1}^{N_e}(\hat{\mu}_i - \mu_e)^2\,. \tag{6}$$

Recall that $\hat{\mu}$ is the sample mean estimation of a single-realization time-series. We can design a model problem which

can be inexpensively simulated an arbitrary number of times and for an arbitrary sample size per realization. At the same time, the samples of such problem need to be autocorrelated as in a statically-stationary turbulence time-series. In particular, we consider a first-order vector autoregressive model to generate samples for two generic variables x and y:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} x_{i-1} \\ y_{i-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{x_i} \\ \varepsilon_{y_i} \end{bmatrix}, \tag{7}$$

where the associated noise samples are chosen to be Gaussian and correlated:

$$[\varepsilon_x, \varepsilon_y]^T \sim \mathscr{N}(\mathbf{0}, \mathbf{C}_\varepsilon), \quad \mathbf{C}_\varepsilon = \begin{bmatrix} a_3^2 & \rho_\varepsilon a_3 b_3 \\ \rho_\varepsilon a_3 b_3 & b_3^2 \end{bmatrix}.$$

To ensure the stationarity condition, the model parameters a_1, a_2, b_1, and b_2 should be chosen such that the eigenvalues of the coefficient matrix in Eq. (7) lie within the unit circle. The experiments to produce Figure 1 are carried out using a_0 to a_2 equal to $0.1\,, 0.5\,, 0.2$ and b_0 to b_2 equal to $0.3\,, 0.4\,, 0.6$. Moreover, $a_3 = 1$, $b_3 = 2$, and $\rho_\varepsilon = 0.5$. Each realization of Eq. (7) starts from a set of random initial samples for x and y taken jointly from $\mathscr{U}[0,1]^2$, and a total of $2n$ samples are generated where the first n samples are discarded for burn-in. Combining moments of x and y, any arbitrary form of statistical terms can be created for which the accuracy of the proposed UQ algorithm can be examined.

First, the sample-estimated uncertainty of the SME of the first-order moment $\langle x \rangle$ is considered. Figure 1(top) shows the PDF (probability density function) of $\hat{\sigma}(\hat{\mu}_x)$ estimated by the ARM-based uncertainty estimator which uses Eq. (2). The PDF contains the empirical σ_e of Eq. (6), and since the mode of the PDF is close to σ_e, the consistency of the ARM-based uncertainty estimator for $\langle x \rangle$ is confirmed.

Next, estimation of uncertainty in the central moments $\langle x'y' \rangle$ and $\langle x'^3 \rangle$ and a nonlinear term $\sqrt{\langle x^2 \rangle \langle y \rangle}$, is considered. For the first two terms, we can compare the performance of the proposed UQ method with the standard method. In the standard method any of the uncertainty estimation techniques in the previous section is applied to the samples of $x'y' = (x - \hat{\mu}_x)(y - \hat{\mu}_y)$ and $x'^3 = (x - \hat{\mu}_x)^3$. This way, the uncertainty in $\hat{\mu}_x$ and $\hat{\mu}_y$ is ignored. In contrast, the proposed UQ algorithm requires decomposing the SME of $\langle x'y' \rangle$ as $\langle x'y' \rangle \approx \mu_{xy} - \mu_x \mu_y + \varepsilon_{\hat{\mu}_{x'y'}}$, where the uncertainty term is,

$$\begin{aligned} \varepsilon_{\hat{\mu}_{x'y'}} \sim (\mathscr{N}(0, \sigma^2_{\hat{\mu}_{xy}}) &- \mathscr{N}(0, \sigma^2_{\hat{\mu}_x}) \mathscr{N}(0, \sigma^2_{\hat{\mu}_y}) \\ &- \mu_x \mathscr{N}(0, \sigma^2_{\hat{\mu}_y}) - \mu_y \mathscr{N}(0, \sigma^2_{\hat{\mu}_x})) \,. \end{aligned} \tag{8}$$

Therefore, $\hat{\boldsymbol{\mu}} = \{\hat{\mu}_x, \hat{\mu}_y, \hat{\mu}_{xy}\}$ in Eq. (3). Having estimated elements of $\boldsymbol{\Sigma}$, expression (8) can be evaluated at any Monte Carlo sample taken from the multivariate Gaussian distribution (3). The result of subtraction/addition of correlated Gaussian distributions is still Gaussian. But multiplication of correlated Gaussian distributions as for instance in Eq. (8) is not Gaussian. Therefore, the uncertainty in the sample estimation of a compound term such as $\langle x'y' \rangle$ can, in general, be non-Gaussian. Although it may happen that the nonlinear term in Eq. (8) is much smaller than the others and hence the resulting distribution of $\varepsilon_{\hat{\mu}_{x'y'}}$ remains Gaussian. This means that the distribution of the uncertainty of a compound statistical term depends on the problem at hand and has to be evaluated a-posteriori. Similarly, $\langle x'^3 \rangle$ can be estimated from $\langle x'^3 \rangle \approx \mu_{x^3} + 2\mu_x^3 - 3\mu_x \mu_{x^2} + \varepsilon_{\hat{\mu}_{x'^3}}$ with the associated uncertainty expressed by,

$$\begin{aligned} \varepsilon_{\hat{\mu}_{x'^3}} = \varepsilon_{\hat{\mu}_{x^3}} &+ 2\varepsilon^3_{\hat{\mu}_x} - 3\varepsilon_{\hat{\mu}_x} \varepsilon_{\hat{\mu}_{x^2}} + 6\mu_x^2 \varepsilon_{\hat{\mu}_x} + 6\mu_x \varepsilon^2_{\hat{\mu}_x} - \\ &- 3\left(\mu_x \varepsilon_{\hat{\mu}_{x^2}} + \mu_{x^2} \varepsilon_{\hat{\mu}_x}\right), \end{aligned} \tag{9}$$

with $\hat{\boldsymbol{\mu}} = [\hat{\mu}_x, \hat{\mu}_{x^2}, \hat{\mu}_{x^3}]^T$. The two middle plots in Figure 1 represent the PDF of the uncertainty in $\langle x'y' \rangle$ and $\langle x'^3 \rangle$ estimated by the proposed and standard UQ methods. For $\langle x'y' \rangle$ both UQ techniques lead to similar PDFs which include the empirical uncertainty estimate. However, for $\langle x'^3 \rangle$, only the proposed algorithm can accurately estimate the sample uncertainty. The fact that the standard method is only accurate for one of these considered moments is problem-dependent, but the main conclusion is that the proposed algorithm is reliable for estimating uncertainty in various moments using single realization data.

Aside from potentially inaccurate estimates, the standard UQ method cannot be applied to many compound statistical terms, for instance $\sqrt{\langle x^2 \rangle \langle y \rangle}$. There is no way to evaluate this term by applying only one SME unless the term is replaced by, for instance, $\langle \sqrt{x^2 y} \rangle$, which is obviously a different statistical term. In contrast, the proposed method requires numerically evaluating $\sqrt{\langle x^2 \rangle \langle y \rangle} \approx \sqrt{(\mu_{x^2} + \varepsilon_{\hat{\mu}_{x^2}})(\mu_y + \varepsilon_{\hat{\mu}_y})}$ for a sufficiently large number of MC samples. This is straightforward, after accurately estimating $\mathrm{cov}(\hat{\mu}_{x^2}, \hat{\mu}_y)$. For this term, the bottom plot in Figure 1 confirms the accuracy of the uncertainty estimated by the proposed method.

APPLICATION TO TURBULENT FLOWS

The proposed UQ algorithm is general and can be applied to any turbulence statistics. Here, the application to the time series obtained from two canonical wall-bounded turbulent flows, the flow over a periodic hill and turbulent pipe flow, is considered. For these flows, there are respectively one and two spatially homogeneous directions at which periodic conditions are applied at each time step during simulation. Representing spatially-averaged quantities by an overbar, a central moment, for instance $\langle \overline{u}'\overline{v}' \rangle$, can be approximately decomposed as $\langle \overline{uv} - \overline{u}\,\overline{v} \rangle = \langle \overline{uv} \rangle - \langle \overline{u}\,\overline{v} \rangle$ which is not, in general, equal to the standard decomposition $\langle \overline{uv} \rangle - \langle \overline{u} \rangle \langle \overline{v} \rangle$, noting that the covariance (in time) between $\overline{u}$ and $\overline{v}$ is not generally zero. This simplification leads to reducing the number of primitive SMEs and hence the size of $\boldsymbol{\Sigma}$ when computing the uncertainty in central moments. For the considered flow cases, this approximation is valid because of having relatively long periodic directions with large number of grid points which result in low uncertainty in spatially-averaged quantities. Note that the standard decomposition can also be treated with the proposed approach which is however not shown here due to brevity.

In all cases discussed below, the uncertainty in the SMEs of the primitive variables is estimated by an ARM method, see Oliver *et al.* (2014); Xavier *et al.* (2022). The steps to estimate uncertainty are the same as those explained in the previous sections and applied to the model problem. A main characteristic of turbulence signals is, however, that the autocorrelations as well as cross-correlations decay much slower than the model problem considered above. This reflects the long history effects in turbulent flows. Moreover, such correlations depend on the variables and spatial location. Thus, careful monitoring of the modeling of the autocorrelations and cross-covariances when using Eqs. (2) and (4), is essential.

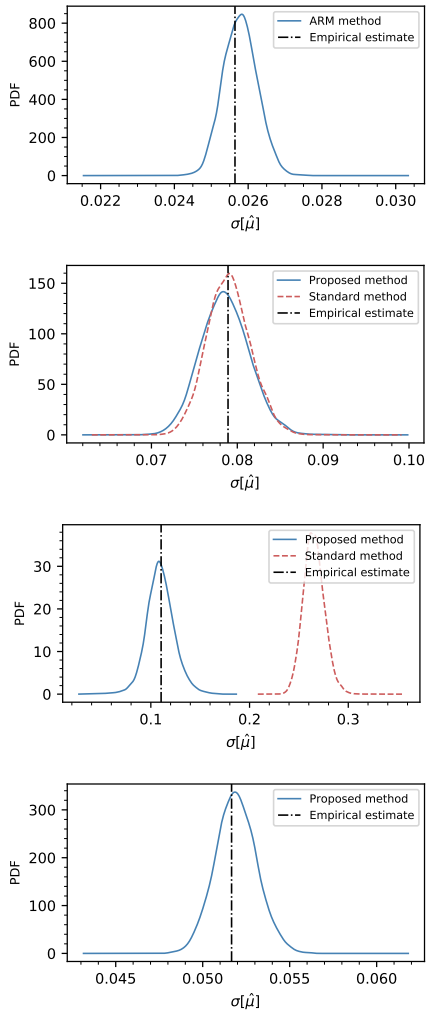

Figure 1. From top to bottom: PDFs of sample-estimated uncertainty of $\langle x \rangle$, $\langle x'y' \rangle$, $\langle x'^3 \rangle$, and $\sqrt{\langle x^2 \rangle \langle y \rangle}$ obtained from the proposed and standard methods compared to the associated empirical uncertainty by Eq. (6) (vertical line). The uncertainties are computed based on $N_e = 10^4$ repetitions of system (7) with sample size $n = 50000$ per realization. The uncertainty in the primitive SMEs (diagonal elements of $\mathbf{\Sigma}$) are estimated by the ARM-based estimator. The off-diagonal elements of $\mathbf{\Sigma}$ are estimated from Eq. (4) using modeled $\gamma_{ij}(\pm k)$.

Flow Over a Periodic Hill

The periodic-hill case is chosen because of its complexity owing to the inherent physical uncertainty in the separated region behind the first hill. The bulk Reynolds number is $Re_b = U_b h/\nu = 2800$ where U_b is the bulk velocity, h is the height of the hill, and ν is the kinematic viscosity. The simulation is performed using the high-order spectral-element solver, Nek5000 (Fischer *et al.*, 2008), based on 9^{th}-order polynomial basis approximations per element in each of the spatial directions. Figure 2 shows the 95% confidence intervals for the time-averaged streamwise velocity $\langle \overline{u} \rangle$ and the Reynolds-stress component $\langle \overline{u}'\overline{u}' \rangle$ at three streamwise locations, $x/h = 2$, 4, and 6, for averaging over 44 flow-through times. Reference data by Breuer *et al.* (2009) are also included. Interestingly, the region with the separated flow i.e. $x/h = 2$ has a lower uncertainty in $\langle u \rangle$ compared to the downstream locations. This is because the autoregressive model can exactly model finite temporal correlations embedded within the time series of the

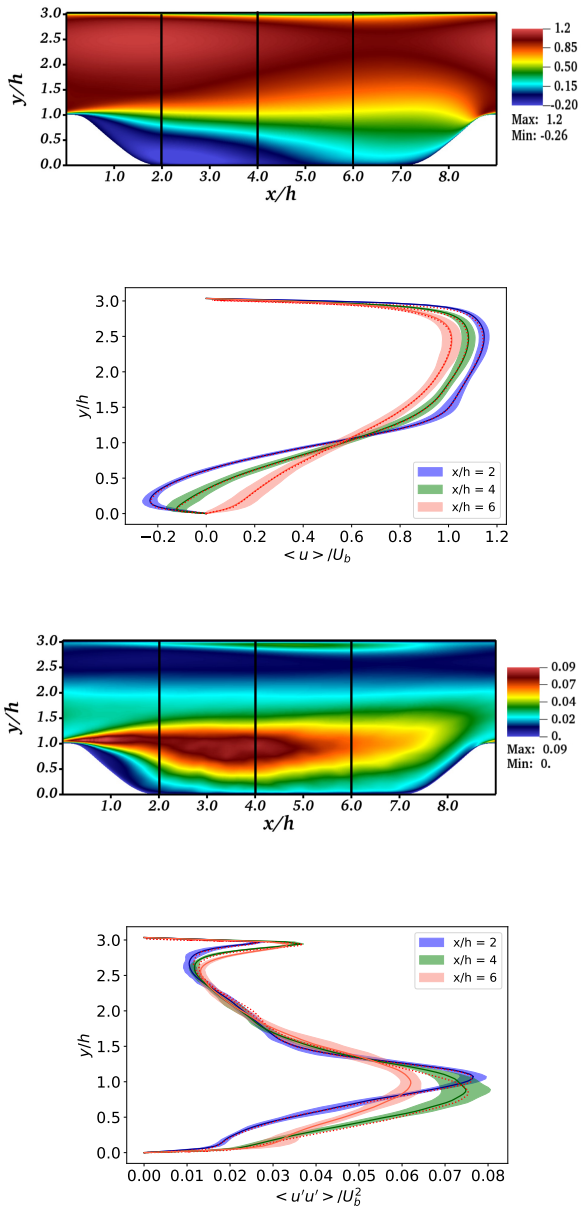

Figure 2. Contours and profiles of $\langle \overline{u} \rangle / U_b$ and $\langle \overline{u}'\overline{u}' \rangle / U_b^2$ of the flow over a periodic hill at $Re_b = 2800$. The profiles are taken at $x/h = 2$, 4, 6 (vertical lines in the contour plots) where solid lines represent the expected value and shaded areas show 95% confidence intervals (CIs). Note that the CIs of $\langle u \rangle / U_b$ are multiplied by 3 for better visibility. The profiles with the dotted lines represent the DNS of Breuer *et al.* (2009).

large separation bubble occurring behind the first hill. The uncertainty in the high-order statistics also reflects this observation. At each of the three streamwise locations, the largest uncertainty in $\langle \overline{u}'\overline{u}' \rangle$ is observed at the peak near the bottom wall. This location is where the shear flow originating from the top of the first hill is developed.

Turbulent Pipe Flow

The second case we considered is turbulent flow in a smooth straight pipe of circular cross-section. The direct numerical simulation (DNS) of the Navier–Stokes equations is conducted using the open-source code "OPENPIPEFLOW",

Table 1. Grid resolution parameters for DNS of turbulent pipe flow at $Re_b = 5300$.

$Nr \times N_\theta \times N_z$	Δr^+	$\Delta(R\theta)^+$	Δz^+
$192 \times 256 \times 1024$	(0.02-2)	4.4	5.5

which is a specific solver developed for pipe flows using Fourier-finite-difference methods (Willis, 2017). The bulk Reynolds number $Re_b = 2U_bR/\nu$ is chosen to be 5300, and the corresponding friction Reynolds number ($Re_\tau \equiv \langle\overline{u}_\tau\rangle R/\nu$ with $\langle\overline{u}_\tau\rangle = (\langle\overline{\tau}_w\rangle/\rho)^{1/2}$ the friction velocity) is 181.95, where ρ, ν, and R are the fluid density, kinematic viscosity, and pipe radius, respectively. In what follows, the radial, azimuthal, and axial directions are represented by r, θ, and z, respectively, and $\bar{\cdot}$ denotes averaging over θ and z. The pipe length L_z/R is 10π and the details of the spatial resolutions are summarized in Table 1. Note that this low Reynolds number is chosen here only for illustration purposes, and we have already obtained data at higher Reynolds numbers up to $Re_\tau = 5200$ (Yao *et al.*, 2021), where the current UQ analyses are applied as well.

Figure 3 shows the the 95% confidence intervals for the first- and second-order velocity moments. All the quantities are inner scaled using the time-averaged wall friction velocity $\langle\overline{u}_\tau\rangle$ and the viscous length scale $\nu/\langle\overline{u}_\tau\rangle$. One of the strong points of the proposed UQ approach is the possibility of accurately including the uncertainty in moments of wall-friction velocity $\overline{u}_\tau$ when computing uncertainty in inner-scaled quantities. For instance, an inner-scaled Reynolds-stress component reads as,

$$\langle\overline{u}_i'\overline{u}_j'\rangle^+ = \frac{\langle\overline{u}_i'\overline{u}_j'\rangle}{\langle\overline{u}_\tau\rangle^2} \approx \frac{\mu_{\overline{u}_i'\overline{u}_j'} + \varepsilon_{\hat{\mu}_{\overline{u}_i'\overline{u}_j'}}}{(\mu_{\overline{u}_\tau} + \varepsilon_{\hat{\mu}_{\overline{u}_\tau}})^2}\,. \tag{10}$$

Having $\hat{\boldsymbol{\mu}} = \{\hat{\mu}_{\overline{u}_i'\overline{u}_j'}, \hat{\mu}_{\overline{u}_\tau}\}$ and associated covariance matrix $\boldsymbol{\Sigma}$ estimated, the right-hand-side of the above expression can be evaluated. As expected, in Figure 3 the uncertainty for $\langle\overline{u}_z\rangle^+$ is large near the pipe center. In addition, it turns out that compared to other Reynolds-stress components, the uncertainty is relatively large for $\langle\overline{u}_z'\overline{u}_z'\rangle^+$, particularly near the peak ($y^+ \approx 15$). Also note that the uncertainty varies between the Reynolds-stress components.

In a similar way, the uncertainty in higher-order velocity moments and any compound term can be estimated. Figure 4 illustrates the uncertainty in the profiles of skewness$(\overline{u}_z) = \langle\overline{u}_z'^3\rangle/\langle\overline{u}_z'^2\rangle^{3/2}$ and kurtosis$(\overline{u}_z) = \langle\overline{u}_z'^4\rangle/\langle\overline{u}_z'^2\rangle^2$. As expected, by increasing the order of a central moment, the uncertainty can increase, compare the uncertainties in kurtosis$(\overline{u}_z)$ and the Reynolds-stress components. In fact, for $y^+ \lesssim 3$, the uncertainty for kurtosis$(\overline{u}_z)$ is considerable. For both terms in Figure 4, an increase in the uncertainty is captured at $y^+ \approx 140$. The contribution of each primitive moment to the uncertainty of a compound term can be quantified in the future using elaborate UQ techniques.

The uncertainty in various budget terms in the transport equations of the Reynolds stresses and turbulent kinetic energy can also be computed. The transport equation for the Reynolds-stress component $\langle\overline{u}_i'\overline{u}_j'\rangle$ in a periodic pipe can be written as,

$$\frac{D\langle\overline{u}_i'\overline{u}_j'\rangle}{Dt} = P_{ij} + TD_{ij} + PD_{ij} + PS_{ij} + VD_{ij} + D_{ij}, \tag{11}$$

for $i, j = 1, 2, 3$, where the successive terms on the right-hand-side of this equation denote production (P), turbulent diffusion (TD), pressure diffusion (PD), pressure strain (PS), viscous diffusion (VD), and dissipation (D) of $\langle\overline{u}_i'\overline{u}_j'\rangle$. The definition of these terms can be found, for instance in Moser & Moin (1984). Estimating the uncertainty in the budget terms is straightforward after decomposing each term into the moments to which the UQ estimators described above can be applied. The only point is about terms having spatial derivatives of moments. For the specific case of turbulent pipe flow, where samples are averaged over z and θ before time-averaging, estimation of uncertainty in $\mathrm{d}\langle\mathbf{M}\rangle/\mathrm{d}r$ is straightforward, noting the differentiation in r can commute with the time-averaging. Therefore, it is enough to collect time samples of elements $\mathbf{M}$ during the simulation. This method has the benefit of retaining the numerical accuracy of the spatial derivatives that is important especially in higher-order flow solvers. An alternative would be extending $\boldsymbol{\Sigma}$ to account for covariances in space (associated to the points in a stencil used for computing spatial derivatives) between primitive SMEs. Figure 5 shows the uncertainty in various terms in Eq. (11) for turbulent kinetic energy (TKE) and $\langle\overline{u}_r'\overline{u}_z'\rangle$. For the latter, the pressure strain and diffusion exhibit larger uncertainty compared to other budgets especially at small y^+. At these locations, the expected value of the mentioned terms also deviate more from the DNS of El Khoury *et al.* (2013).

CONCLUSIONS

A framework is introduced for accurately estimating time-average uncertainty in sample-estimated statistics of turbulent flow simulations. A set of guidelines connected to the turbulence physics is introduced to adjust the hyperparameters which control the accuracy of various uncertainty estimators. For quantifying uncertainties in high-order and general turbulence statistics, a novel UQ algorithm is proposed which takes into account autocorrelation and cross-covariances in turbulence time-series and associated sample-mean estimators. The approach shows an excellent accuracy when its sample uncertainty estimations are validated against the ensemble empirical uncertainties for a model problem. The algorithm is demonstrated to be general and accurate for various types of statistics, some of which cannot be dealt with using standard uncertainty estimation methods. To illustrate the application in turbulent flow simulations, DNS of two canonical wall-bounded turbulent flows are considered. A next development step will include transferring the workflow to the in-situ framework of Gscheidle *et al.* (2022). This will result in a set of efficient and reliable tools for monitoring convergence of various statistics in large-scale simulations of turbulent flows independent of the flow solver.

ACKNOWLEDGMENTS

This work has been supported by: i) EXCELLERAT project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823691 (SR, PS), and ii) TTU Distinguished Chair funding (JY, FH). The flow simulations were performed on the resources provided by the Swedish National Infrastructure for Computing (SNIC) at PDC (KTH) and Frontera at Texas Advanced Computing Center.

REFERENCES

Benedict, L. H. & Gould, R. D. 1996 Towards better uncertainty estimates for turbulence statistics. *Exp. Fluids* **22** (2), 129–136.

Breuer, M., Peller, N., Rapp, Ch. & Manhart, M. 2009 Flow over periodic hills – numerical and experimental study in

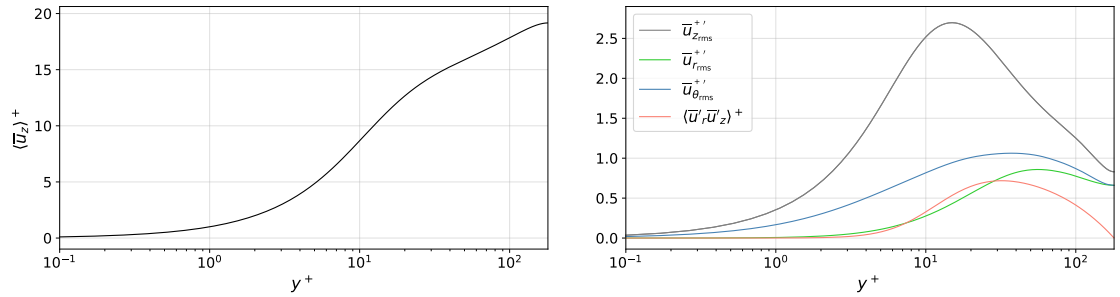

Figure 3. Profiles of inner-scaled (left) mean axial velocity and (right) Reynolds-stress components of turbulent pipe at $Re_b = 5300$. Solid lines represent the expected values and shaded areas show 95% CIs due to time-average uncertainties. Note that the CIs are multiplied by 5 for better visibility. The uncertainty in $\langle \overline{u}_\tau \rangle$ is ignored for $y^+ = \langle \overline{u}_\tau \rangle (R-r)/\nu$.

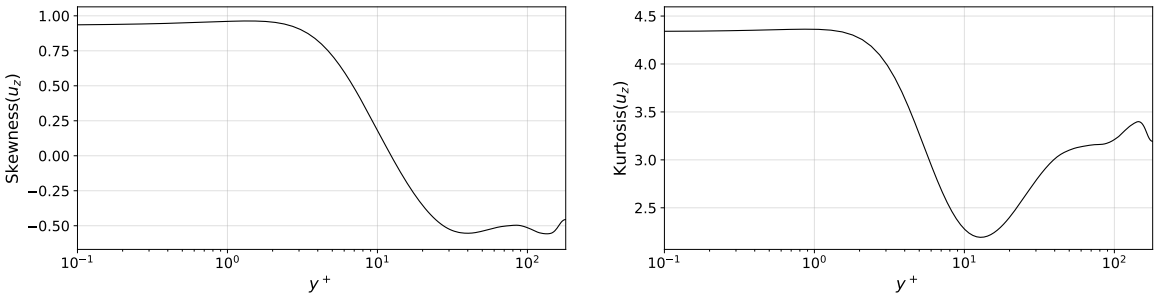

Figure 4. Profiles of (left) skewness and (right) kurtosis of the streamwise velocity $\overline{u}_z$ of the turbulent pipe flow at $Re_b = 5300$.

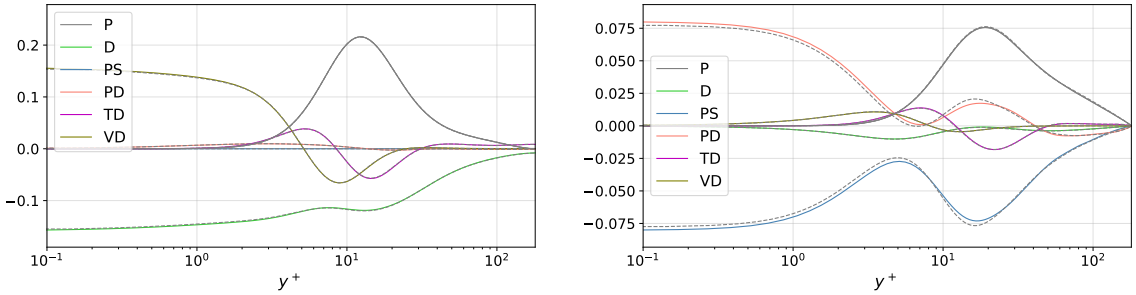

Figure 5. Inner-scaled budget of (left) turbulent kinetic energy and (right) $\langle \overline{u}'_r \overline{u}'_z \rangle$. The 95% CIs of the turbulent kinetic energy are multiplied by 5 for better visibility. The dashed lines represent the DNS of El Khoury *et al.* (2013).

a wide range of Reynolds numbers. *Comput. Fluids* **38** (2), 433–457.

El Khoury, G. K., Schlatter, P., Noorani, A., Fischer, P. F., Brethouwer, G. & Johansson, A. V. 2013 Direct numerical simulation of turbulent pipe flow at moderately high Reynolds numbers. *Flow Turbul. Combust.* **91** (3), 475–495.

Fischer, P. F., Lottes, J. W. & Kerkemeier, S. G. 2008 NEK5000: Open source spectral element CFD solver. Available at: http://nek5000.mcs.anl.gov.

Gscheidle, C., Rezaeiravesh, S., Garcke, J. & Schlatter, P. 2022 In-situ estimation of time-averaging uncertainties in turbulent flow simulations. In *Workshop MMCP, HPC Asia 2022, Jan 12-14, Kobe, Japan*.

Hoyas, S. & Jiménez, J. 2008 Reynolds number effects on the Reynolds-stress budgets in turbulent channels. *Phys. Fluids* **20** (10), 101511.

Lee, M. K. & Moser, R. D. 2015 Direct numerical simulation of turbulent channel flow up to $Re_\tau \approx 5200$. *J. Fluid Mech.* **774**, 395–415.

Moser, R. D. & Moin, P. 1984 Direct numerical simulation of curved turbulent channel flow. NASA Technical Memorandum 85974.

Oliver, T. A., Malaya, N., Ulerich, R. & Moser, R. D. 2014 Estimating uncertainties in statistics computed from direct numerical simulation. *Phys. Fluids* **26** (3), 035101.

Rezaeiravesh, S., Vinuesa, R. & Schlatter, P. 2021 UQit: A Python package for UQ in CFD. *JOSS* **6** (60), 2871.

Rezaeiravesh, S., Vinuesa, R. & Schlatter, P. 2022 An uncertainty-quantification framework for assessing accuracy, sensitivity, and robustness in computational fluid dynamics. *J. Comput. Sci.* **62**, 101688.

Russo, S. & Luchini, P. 2017 A fast algorithm for the estimation of statistical error in DNS (or experimental) time averages. *J. Comput. Phys.* **347**, 328 – 340.

Smith, R. C. 2013 *Uncertainty Quantification: Theory, Implementation, and Applications*. Philadelphia: SIAM.

Willis, A. P. 2017 The Openpipeflow Navier–Stokes solver. *SoftwareX* **6**, 124–127.

Xavier, D., Rezaeiravesh, S., Vinuesa, R. & Schlatter, P. 2022 Reliable quantification of time-averaging uncertainties in turbulent flow simulations. *To be submitted* .

Yao, J., Schlatter, P., Rezaeiravesh, S. & Hussain, F. 2021 Direct numerical simulation of turbulent pipe flow up to Re_τ=5200. *Bulletin of the American Physical Society* .