Stochastic estimation as a statistical tool for approximating turbulent conditional averages

G. J. Brereton
Department of Mechanical Engineering and Applied Mechanics, The University of Michigan, Ann Arbor, Michigan 48109

In this paper, extensions to conventional stochastic estimation techniques are presented, whereby uncertainties in individual estimates may be deduced. Test applications to time series of velocity measurements in a turbulent boundary layer confirm the fidelity of the uncertainty estimation procedure and illustrate how the optimal choice of stochastic estimation model can be strongly dependent on the event upon which the average is conditioned. They also demonstrate how stochastic estimations may be refined to yield more accurate descriptions of particular coherent motions, and how they can reveal the existence of rare events, different in statistical character to their more frequent counterparts, which might otherwise be undetected by conventional stochastic estimation.

I. INTRODUCTION

Stochastic estimation has recently been introduced as a procedure for approximating turbulent conditional averages and has been employed to identify and describe coherent motions of turbulent flow. Its applications have ranged from characterizing the conditional eddies of isotropic turbulence as vortex rings (Adrian[1,2]) to describing the spatiotemporal evolution of the turbulent velocity field during conditionally detected events in flat-plate turbulent boundary layers (Guezennec[3] and Choi and Guezennec[4]). Related applications of stochastic estimates include descriptions of the dominant structures in jet mixing layers (Cole et al.[5]) and characterization of the turbulent motions of turbulent pipe flow (Hassan et al.[6]) as large eddies of azimuthal vorticity. Evaluations of stochastic estimation as an alternative experimental method to conditional averaging of single- and multipoint measurements have been undertaken in a range of turbulent flows by Adrian et al.[7] In addition, the technique has been proposed as an efficient means of processing three-dimensional imaged flow fields in order to make quantitative inferences about their momentary character (Adrian[8]), and has been implemented for the case of two-dimensional fields in fully developed channel flow by Nishino and Kasagi.[9]

Stochastic estimation has been used for analysis of three-dimensional databases describing instantaneous velocity and pressure fields in low Reynolds number turbulent flows, for which numerical simulation of the Navier–Stokes equations is possible. When postprocessing such databases, the interpretation of vast quantities of spatiotemporal information is poorly suited to conventional tools for turbulent structure identification and techniques such as stochastic estimation appear to hold great promise. In this application, Adrian and Moin[10] have used linear stochastic estimation to identify the flow features that contribute most strongly to generation of Reynolds stress in homogeneous shear flow, while Moin et al.[11] have characterized the motions that play the same roles in turbulent channel flow.

The stochastic estimation techniques employed in the aforementioned studies are based on the premise that statistical descriptions of events of interest (which may be infrequent or of high dimension) may be obtained from data sets with greater ease and better statistical convergence if those descriptions are modeled on the behavior of all possible events contained within the data set. The models that have been employed are linear, quadratic, or cubic polynomials, the modeling parameters of which are obtained by least-squares fitting. This form of stochastic estimation procedure appears to be well suited to characterizing rare events in fields, which may be modeled reliably by least-squares fits of polynomial conditional-average models, when the statistics of those events are closely related to those of more frequent events. The purposes of this paper are to develop guidelines for ascertaining when statistical events conform to these requirements through application of refined forms of stochastic estimation to characterizing features of boundary-layer turbulence.

II. THEORETICAL BASIS FOR STOCHASTIC ESTIMATION

As described in detail by Adrian et al.[7] and Guezennec,[3] stochastic estimation is a technique for making an estimation of the conditional average of some quantity g given the event $\mathbf{E}$. Following the notation of Adrian et al.[7] this conditional average is expressed as $\langle g|\mathbf{E}\rangle$, defined as

$$\langle g|\mathbf{E}\rangle = \int \frac{f(g,\mathbf{E})}{f(\mathbf{E})} g\, dg, \tag{1}$$

which employs statistical information about all possible values of the quantity g. Here, $f(\mathbf{E})$ is the probability density function for the event vector $\mathbf{E}$, upon which the occurrence of g is conditioned, and $f(g,\mathbf{E})$ is their joint probability density function. The quantity g is usually taken as the component of g deviatoric from its temporal or spatial

mean. The integral expression in (1) is typically modeled in the general form of a series of positive powers of $\mathbf{E}(\mathbf{x},t)$, as

$$\langle g_i(\mathbf{x}+\mathbf{r},t+\tau) \,|\, \mathbf{E}(\mathbf{x},t)\rangle$$

$$= A_i(\mathbf{r},\tau) + B_{ij}(\mathbf{r},\tau)E_j(\mathbf{x},t) + C_{ijk}(\mathbf{r},\tau)E_j(\mathbf{x},t)E_k(\mathbf{x},t)$$

$$+ D_{ijkl}(\mathbf{r},\tau)E_j(\mathbf{x},t)E_k(\mathbf{x},t)E_l(\mathbf{x},t) + \cdots . \qquad (2)$$

For any values of temporal and spatial displacements $\mathbf{r}$ and τ, this model of $\langle g_i(\mathbf{x}+\mathbf{r},t+\tau) \,|\, \mathbf{E}(\mathbf{x},t)\rangle$ has obvious physical plausibility as a Taylor series expansion about $g_i(\mathbf{x},t)$ in $\mathbf{E}(\mathbf{x},t)$ space, providing the event vector $\mathbf{E}(\mathbf{x},t)$ (or its deviation from its mean) may be treated as differential in scale. When it is of larger scale, the model is effectively a polynomial of positive powers of $\mathbf{E}(\mathbf{x},t)$, which may capture the physical attributes of particular turbulent conditional averages with varying degrees of success, depending on the displacements $\mathbf{r}$ and τ, and depending on the statistical nature of the flow in question. A motivating factor for modeling of this kind is that when the distributions of the data and event vectors about their respective means are

jointly normal, the optimal estimate (in a least-squares sense) of the data vector is a linear scaling of the event vector.[12] Thus the inclusion of nonlinear terms in the stochastic estimation model may be viewed as recognition that real distributions of turbulent data and event vectors are not jointly normal, in which case the least-squares procedure may not be the most appropriate minimization. When incorporated in practical estimation procedures, the power series is truncated at low orders. For example, in the studies of Adrian et al.,[7] all terms beyond the linear one are discarded; the implementation of Guezennec[3] includes terms as high as the quadratic and cubic.

The coefficients of the stochastic model are deduced by a maximum likelihood estimation. Given a probability density function for the deviation of data from the conditional average model, an estimate is made of the set of coefficients for which it is most likely that the data existed. Following standard developments of the theory of maximum likelihood estimation,[13] it can be shown that the model coefficients may be found by a minimization of the function

$$\sum_{n=1}^{N} \rho\left(\frac{g_i(n) - A_i(\mathbf{r},\tau) - B_{ij}(\mathbf{r},\tau)E_j(n) - C_{ijk}(\mathbf{r},\tau)E_j(n)E_k(n) - \cdots}{\sigma(n)}\right). \qquad (3)$$

The minimization is carried out over all members of an ensemble of N data $g_i(n)$, evaluated at $\mathbf{x}+\mathbf{r}$, $t+\tau$, for the N events $\mathbf{E}(n)$, evaluated at $\mathbf{x},t$. Here, ρ is the negative logarithm of the probability density function for the deviation of data from the conditional average model and $\sigma(n)$ is the standard deviation of each $g_i(n)$, $\mathbf{E}(n)$ measurement pair. Typically, $\sigma(n)$ is chosen as uniform over all members of $g_i(n)$ and $\mathbf{E}(n)$, with ρ assumed to take a Gaussian distribution for all $\mathbf{r}$ and τ. This particular set of assumptions corresponds to the least-squares minimization procedure, in which case $A_i(\mathbf{r},\tau)$ is zero when g_i is the component deviatoric from its mean.

The least-squares minimization procedure corresponds to a maximum likelihood estimate of the stochastic model parameters if the deviations of data from the estimate are independent and follow a normal distribution. Although normalcy is often a useful approximation to the distribution of turbulent measures about their means, it is rarely an accurate model. The condition of joint normalcy between turbulent data and event measures can be more difficult to satisfy—many turbulent probability density functions exhibit skewness and the tails of these distributions are often very different in form from Gaussian ones.[14] This point is central to the issue of accuracy in stochastic estimation, since it is often in estimation of rare events within the tails of distributions that methods that perform more reliably than conditional averaging appear to be most useful. Another difficulty that arises in deployment of any maximum likelihood estimation is the presumption that individual

deviations of data from the optimal fit (the stochastic estimate) to those data are independent. The recognition of coherence in turbulent motions over a wide range of scales implies that deviations of data from low-order models are likely to retain some degree of organization and that such deviations are therefore not strictly independent. This shortcoming arises because decomposition of fields into simply a mean and a turbulent component is not entirely satisfactory, and proposals such as decomposition by class of phase-space structure (Farge[15]) appear to offer the hope of retaining dynamically meaningful information in which deviations from a class mean are more likely to be independent. However, until better methods are available, the issue of independence of individual deviations of data from means seems unavoidable in providing simple statistical characterizations of turbulent motion. What is also of concern is how one should make a judicious choice of an appropriate stochastic model and maximum likelihood procedure for the accurate estimation of data that follow nonnormal distributions. An approach that provides guidance in making this choice, and in assessing the degree of confidence to be placed in an estimate, is presented in this paper.

III. ACCURACY ASSESSMENT IN STOCHASTIC ESTIMATION

In the applications of stochastic estimation discussed in the Introduction, the stochastic estimate is modeled as a smooth function of the event vector. The coefficients of the

2047   Phys. Fluids A, Vol. 4, No. 9, September 1992

G. J. Brereton   2047

stochastic estimate are evaluated independently for different displacements in time and space. However, an estimate of a conditional average at a given event could be modeled just as plausibly as a smooth function over small temporal or spatial displacements as it could over displacements in event vector space. This two-dimensional smoothness may be exploited by comparing estimates for the same time/position and event, based on modeling in event vector space and on temporal/spatial modeling. Close agreement between estimates obtained from modeling in different dimensions would be consistent with accuracy in both (i) models of the variation in time/space of the stochastic estimate at a given event and (ii) models of the dependence of the stochastic estimate on the event vector at a given time and position. When the displacements in time or position are small and readily modeled by simple functions, the comparison of estimates for the same time/position and event provides a test that is indicative of the adequacy of the event vector model at those events used to make the comparison. Accurate estimates based on faithful event vector models applied at adjacent times or positions would be consistent with a smooth variation of the estimate with time or space and would lead to small discrepancies in the comparison. Conversely, stochastic models that described adjacent positions in event vector space with quite differing degrees of accuracy would produce erratic variations of conditional-average estimates with time/space, revealing appreciable discrepancies in a comparison of the kind proposed. In this way, the smoothness of the estimate with both temporal/spatial variation and variation over the event vector serves as a guide to the adequacy of stochastic modeling.

A simple method of monitoring the smoothness of the stochastic estimate of a conditional average is through comparison of (i) the estimate for a displacement of 2τ (or $2\mathbf{r}$), given the event $\mathbf{E}$; with (ii) the estimate for displacement τ (or $\mathbf{r}$), given the estimate at displacement τ (or $\mathbf{r}$) conditioned on the event $\mathbf{E}$. These single-step and two-step estimates are subsequently referred to as $\langle g|\mathbf{E}\rangle_1$ and $\langle g|\mathbf{E}\rangle_2$. The difference between the two estimates is a convenient indicator of the modeling truncation error for the particular displacement chosen; this technique enjoys widespread use in controlling step size in numerical computations.

For simplicity, we first consider this comparison of estimates as a function of temporal resolution. If the truncation error Δ is taken as

$$\Delta = \langle g|\mathbf{E}\rangle_1 - \langle g|\mathbf{E}\rangle_2 ,$$

and, for displacements τ and $\mathbf{r}$, $\langle g|\mathbf{E}\rangle$ is modeled as

$$\langle g_i(\mathbf{x}+\mathbf{r},t+\tau)|\mathbf{E}(\mathbf{x},t)\rangle = A_i(\mathbf{r},\tau) + B_{ij}(\mathbf{r},\tau)E_j(\mathbf{x},t) + \cdots,$$

then the single time-step conditional average $\langle g|\mathbf{E}\rangle_1$ is simply

$$\langle g_i(\mathbf{x}+\mathbf{r},t+2\tau)|\mathbf{E}(\mathbf{x},t)\rangle$$

$$= A_i(\mathbf{r},2\tau) + B_{ij}(\mathbf{r},2\tau)E_j(\mathbf{x},t) + \cdots . \qquad (4)$$

The two-step conditional average $\langle g|\mathbf{E}\rangle_2$ is written as

$$\langle g_i(\mathbf{x}+\mathbf{r},t+2\tau)|\langle g_i(\mathbf{x}+\mathbf{r},t+\tau)|\mathbf{E}(\mathbf{x},t)\rangle\rangle,$$

and may be modeled by an expansion of $g_i(\mathbf{x}+\mathbf{r},t+2\tau)$ in $\langle g_i(\mathbf{x}+\mathbf{r},t+\tau)|\mathbf{E}(\mathbf{x},t)\rangle$ space, taking the form

$$\langle g_i(\mathbf{x}+\mathbf{r},t+2\tau)|\langle g_i(\mathbf{x}+\mathbf{r},t+\tau)|\mathbf{E}(\mathbf{x},t)\rangle\rangle$$

$$= P_i(\mathbf{r},\tau) + Q_{ij}(\mathbf{r},\tau)[A_j(\mathbf{r},\tau)$$

$$+ B_{jk}(\mathbf{r},\tau)E_k(\mathbf{x},t) + \cdots] + \cdots . \qquad (5)$$

In general, the coefficients P, Q, etc., bear no simple relation to the stochastic estimation coefficients A, B, etc., and must be determined by a second minimization for which an assumption about the distribution of $g_i(\mathbf{x}+\mathbf{r},t+2\tau)$ about its conditional average estimated at $\mathbf{x}+\mathbf{r},t+\tau$ must be made. However, when data and event vectors are the same, as in the time-delayed conditional average of $\langle u(\mathbf{x},t+\tau)|u(\mathbf{x},t)\rangle$, the coefficients of each expansion are identical at each order. The estimate of the truncation error for a displacement in time of 2τ may be expressed in the general form

$$\Delta_i(2\tau) = A_i(\mathbf{r},2\tau) + B_{ij}(\mathbf{r},2\tau)E_j(\mathbf{x},t) + \cdots - P_i(\mathbf{r},\tau)$$

$$- Q_{ij}(\mathbf{r},\tau)[A_j(\mathbf{r},\tau) + B_{jk}(\mathbf{r},\tau)E_k(\mathbf{x},t)$$

$$+ \cdots] - \cdots , \qquad (6)$$

or, in the case of identical data and event vectors,

$$\Delta_i(2\tau) = A_i(\mathbf{r},2\tau) + B_{ij}(\mathbf{r},2\tau)E_j(\mathbf{x},t) + \cdots - A_i(\mathbf{r},\tau)$$

$$- B_{ij}(\mathbf{r},\tau)[A_j(\mathbf{r},\tau) + B_{jk}(\mathbf{r},\tau)$$

$$\times E_k(\mathbf{x},t) + \cdots] - \cdots . \qquad (7)$$

The corresponding truncation error evaluation as a function of spatial resolution is found by formulating the equivalent single-step ($2\mathbf{r}$) and two-step ($\mathbf{r}$) estimates. With appropriate choices of tolerance on Δ and sufficiently small displacements to warrant modeling the spatial/temporal variation as a Taylor series expansion, this method may be used to monitor the region of trustworthiness of the estimate. It may also serve as a guide to the incremental displacements in measurements of g necessary to achieve predetermined levels of trustworthiness. Since the accuracy of the estimate will vary with time, position, and, according to the event upon which it is conditioned, an accuracy check of this kind is a useful accompaniment to stochastic estimation procedures for which confidence limits are to be ascribed.

As an example of the implementation of this procedure, we consider the single-point time-delayed conditional average of the turbulent component of streamwise velocity u, deduced from a time series of 32 768 measurements in a flat-plate turbulent boundary layer ($y/\delta = 0.4$, $\mathrm{Re}_\theta = 3100$). Conditional averages of the time-delayed conditional average of $u'(t)$ computed from this data set are shown in Fig. 1 for a range of different events, plotted against time expressed in viscous units. These averages of $u'(t+\tau)$ were conditioned on $u'(t)$ lying within a window of $\pm 0.1 u'/\sigma$ around the specified level of u', where σ de-
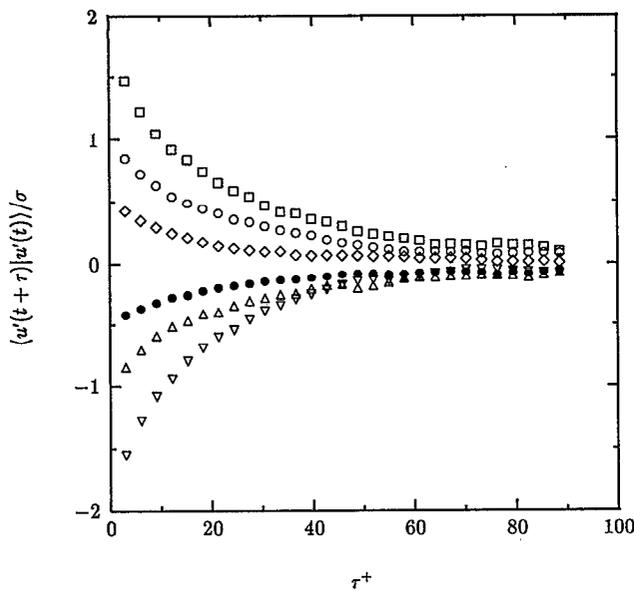
FIG. 1. Time-delayed conditional averages of $u'(t)$; $\lozenge$, conditional average for $u'/\sigma=0.5$; $\bullet$, conditional average for $u'/\sigma=-0.5$; $\bigcirc$, conditional average for $u'/\sigma=1.0$; $\triangle$, conditional average for $u'/\sigma=-1.0$; $\square$, conditional average for $u'/\sigma=2.0$; $\triangledown$, conditional average for $u'/\sigma=-2.0$. The abscissa represents the time delay in viscous units.
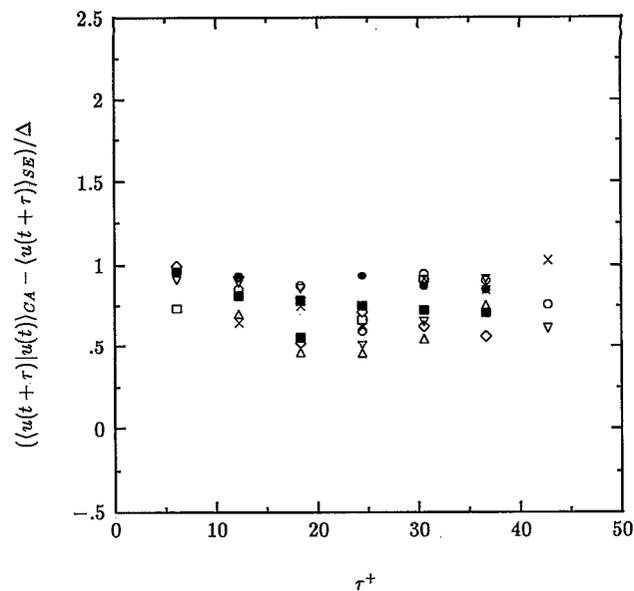
FIG. 2. Comparison of the error estimate Δ with the deviation of the stochastic estimate from the conditional average; $\bigcirc$, linear, $u'/\sigma=1.0$, $\triangle$, linear, $u'/\sigma=1.5$, $\square$, quadratic, $u'/\sigma=1.5$, $\triangledown$, linear, $u'/\sigma=-1.0$, $\lozenge$, linear, $u'/\sigma=-1.5$, $\bullet$, quadratic, $u'/\sigma=-1.5$, $\blacksquare$, cubic, $u'/\sigma=-1.5$, $\times$, quadratic, $u'/\sigma=-2.0$.

notes the standard deviation of u' from its mean. The corresponding stochastic estimate $\langle u(\mathbf{x},t+\tau)|u(\mathbf{x},t)\rangle$ is modeled from (2) as

$$\langle u(\mathbf{x},t+\tau)|u(\mathbf{x},t)\rangle$$
$$=A(0,\tau)+B(0,\tau)u(\mathbf{x},t)+C(0,\tau)u^2(\mathbf{x},t)$$
$$+D(0,\tau)u^3(\mathbf{x},t)+\cdots,$$

and for the case of the linear estimate, the truncation error estimate $\Delta(2\tau)$ is given by

$$\Delta(2\tau)=A(0,2\tau)-A(0,\tau)-A(0,\tau)B(0,\tau)+[B(0,2\tau)$$
$$-B^2(0,\tau)]u(\mathbf{x},t). \qquad (8)$$

Extension of this error estimate for truncation at higher orders is straightforward. If the deviation of data from the conditional average model is assumed to follow a Gaussian distribution, the minimization procedure requires A to be zero, further simplifying evaluation of $\Delta(2\tau)$ as

$$\Delta(2\tau)=[B(0,2\tau)-B^2(0,\tau)]u(\mathbf{x},t). \qquad (9)$$

The estimate of the truncation error in the stochastic model may be compared with the difference between the measured conditional average and the stochastic estimate of $\langle u(\mathbf{x},t+\tau)|u(\mathbf{x},t)\rangle$, in order to gauge the adequacy of the estimated truncation error as an indicator of accuracy. This comparison is made for the case of least-squares stochastic estimation for linear, quadratic, and cubic representations of $\langle u(\mathbf{x},t+\tau)|u(\mathbf{x},t)\rangle$, and is shown in Fig. 2. In this figure, $[\langle u(\mathbf{x},t+\tau)|u(\mathbf{x},t)\rangle_{\text{cond. av.}}-\langle u(\mathbf{x},t+\tau)|u(\mathbf{x},t)\rangle_{\text{stoch. est.}}]/\Delta$ is plotted against time in viscous units for estimates and averages conditioned on $u'(t)=\pm\sigma$, $\pm1.5\sigma$, and $\pm2\sigma$. They are plotted for those

stochastic estimates for which the estimated truncation error exceeded 4% of σ—smaller errors approached the order of the uncertainty in the conditional average and ratios of those errors were not considered meaningful. Beyond time delays of 50 viscous time units (around two integral scales), few reliable error estimates were available from this data set because the conditional average approached zero. While these data do not collapse to a constant value, owing to the approximate nature of the stochastic model and to uncertainty in the conditional averages, very few of these data fall beyond a narrow range from 0.5 to 0.9. The implication of these results is that the estimate of the truncation error can be a reliable indicator of the discrepancy between the stochastic estimate and the conditional average when it is sufficiently greater than an uncertainty threshold, representative of the trustworthiness of the data. It therefore may be used as a guide to the optimal model or fit for refined applications of stochastic estimation, and as a guide to the trustworthiness of individual estimates. This finding is significant since existing techniques of stochastic estimation provide no direct indication of their accuracy—their adequacy has typically been demonstrated by showing agreement between estimates and conditional averages for specific events (Guezennec[3] and Adrian et al.[7]).

The fidelity of the estimated truncation error as a guide to the uncertainty in a stochastic estimate is a consequence of the short (differential) time step between data and the reasonableness of assuming a Gaussian distribution of data about polynomial models for conditional-average estimates in this particular data set. Time steps that can follow the greatest curvature of conditional averages are necessary for concurrent error estimates and so should be of the order of the Taylor microscale of the flow. The reasonableness of
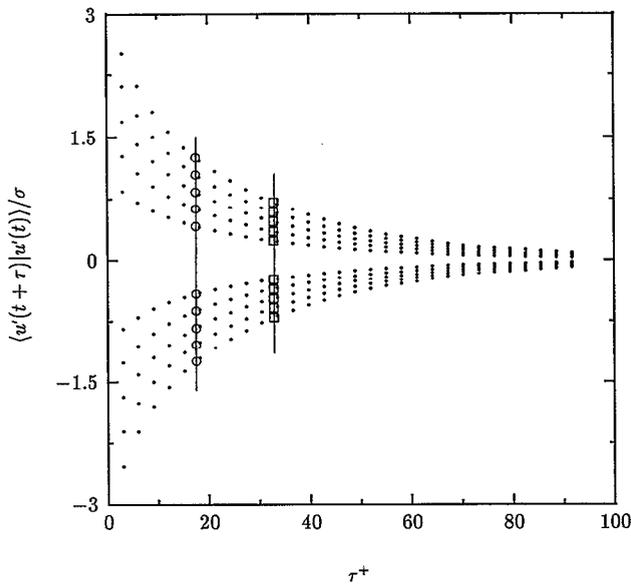
FIG. 3. Least-squares linear stochastic estimates of the time-delayed conditional average of $u'(t)$; $\bigcirc$, estimates of 10% relative uncertainty; $\square$, estimates of 20% relative uncertainty. The estimates are for the events $u'/\sigma = \pm 1$, ± 1.5, ± 2, ± 2.5, and ± 3.
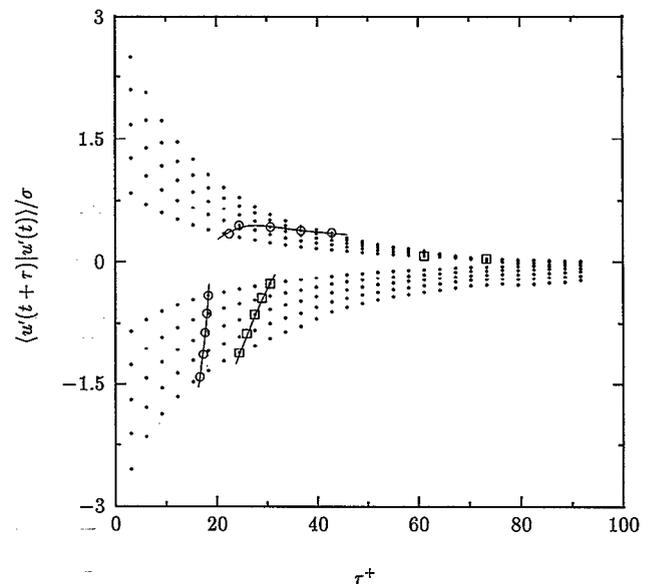

FIG. 4. Least-squares quadratic stochastic estimates of the time-delayed conditional average of $u'(t)$; $\bigcirc$, estimates of 10% relative uncertainty; $\square$, estimates of 20% relative uncertainty. The estimates are for the events $u'/\sigma = \pm 1$, ± 1.5, ± 2, ± 2.5, and ± 3.

employing a Gaussian model is consistent with results for data sets from a number of other turbulent flows. Thus one might expect estimates of the truncation error to serve equally well as indicators of the uncertainty in stochastic estimates in many other turbulent flows. They may also be employed as a guide to refined applications of stochastic techniques in turbulent flow. Several applications are presented in the following sections concerning refinement of the stochastic model, its order, its fit, and the choice of data space to improve eduction of coherent motions. They are, for the test case of a time series of streamwise velocity, measured in a flat-plate turbulent boundary layer, for which the fidelity of the truncation error estimate was demonstrated.

A. Selection of the order of the stochastic estimate

If the estimate of the truncation error is expressed as a fraction of the stochastic estimate for a given event, it represents the relative uncertainty of that estimate, thereby allowing systematic comparisons of uncertainties in estimates made from different event vector models, so that the most favorable model may be selected for a particular estimate. To illustrate this capability, we plot the least-squares stochastic estimate of the time-delayed conditional average of $u'(t)$ using linear, quadratic, and cubic models for the event vector. These estimates are plotted in Figs. 3–5, respectively, together with contours joining points of 10% and 20% relative uncertainties in the estimates. The stochastic estimates are plotted for $u' = \pm \sigma$, $\pm 1.5\sigma$, $\pm 2\sigma$, $\pm 2.5\sigma$, and $\pm 3\sigma$. Since estimates made with the linear model are in direct proportion to the magnitude of the event, and the estimated error contours are vertical lines; the same relative error would be expected for both frequent and rare conditional events. The quadratic and cubic mod-

els contain additional information concerning higher-order moments, such that the extra degrees of freedom permitted in the event vector model might be expected to produce better fits, on average. From inspection of Fig. 4, it is evident that the inclusion of a quadratic term results in significant improvements in the accuracy of the estimate of events for which $u' > \sigma$, but at the expense of reduced accuracy in estimates for $u' < -\sigma$. The addition of a cubic term (Fig. 5) causes little change in the accuracy of esti-
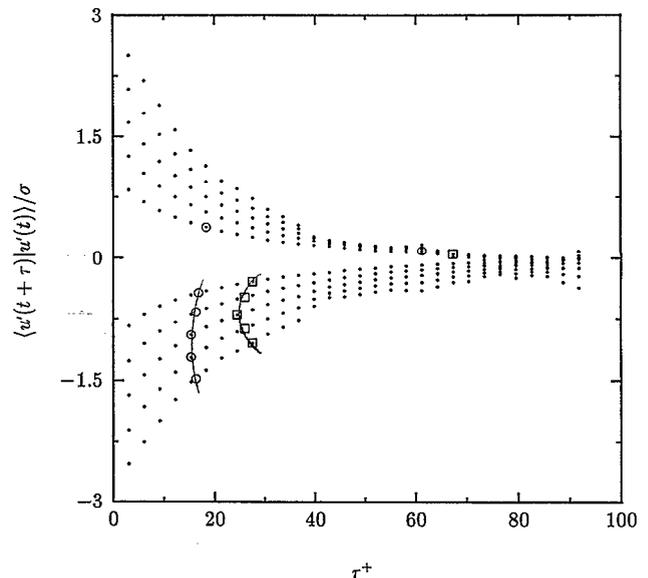

FIG. 5. Least-squares cubic stochastic estimates of the time-delayed conditional average of $u'(t)$; $\bigcirc$, estimates of 10% relative uncertainty; $\square$, estimates of 20% relative uncertainty. The estimates are for the events $u'/\sigma = \pm 1$, ± 1.5, ± 2, ± 2.5, and ± 3.
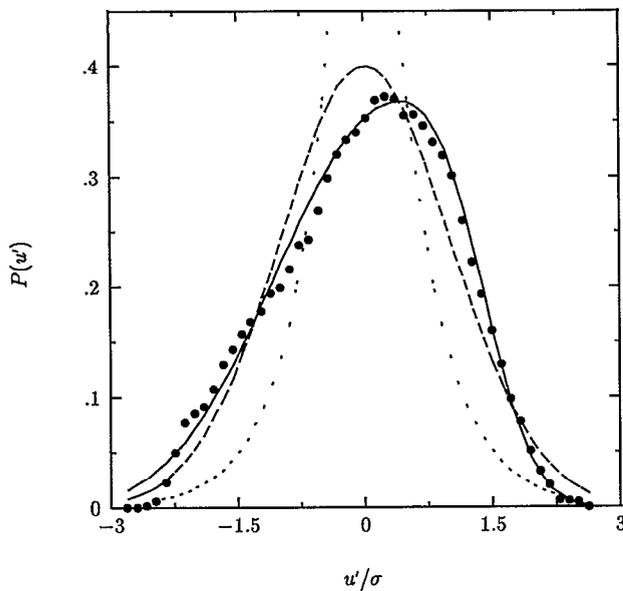
FIG. 6. Comparisons of measured and modeled probability density functions for stationary turbulent boundary-layer velocity data; ●, measurements sorted into bins; ——, distribution modeled in the analytical form of Barndorff-Nielsen;[14] - - -, Gaussian distribution;···, double exponential distribution.
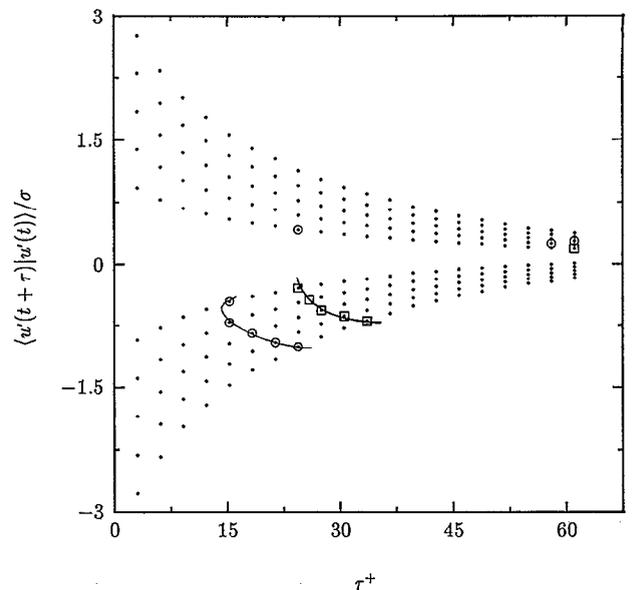
FIG. 7. Quadratic stochastic estimates of the time-delayed conditional average of $u'(t)$ in which model coefficients are deduced by minimizing the sum of the absolute deviation of data from the model; O, estimates of 10% relative uncertainty; □, estimates of 20% relative uncertainty. The estimates are for the events $u'/\sigma = \pm 1$, ± 1.5, ± 2, ± 2.5, and ± 3.

mates for $u' > \sigma$, though there is improvement for $u' < -\sigma$ relative to quadratic estimates of these events. For this data set, quadratic or cubic estimates are preferable for the estimation of events that correspond to large positive u', while a linear estimate yields the best performance of these models for events for which u' is significantly below its mean value. A similar dependence on model order has been noted by Guezennec.[3]

The explanation for these findings lies in the measured probability distribution function of the data shown in Fig. 6 together with a Gaussian distribution for the same standard deviation. For the case of a least-squares fit, the greatest contribution to a model linear in u' is at the maximum of $u'e^{-(u'/\sigma)^2}$, which is found at $u'/\sigma = \pm 1/\sqrt{2}$. The peak contribution to the quadratic term in u' (the maximum of $u'^2 e^{-(u'/\sigma)^2}$) is at $u'/\sigma = \pm 1$. From the distribution of Fig. 6, the contribution at $u'/\sigma = 1$ is significantly larger than that at $u'/\sigma = -1$. The contribution to the cubic term (at $u'/\sigma = \sqrt{3/2}$) is also greater than that at $-\sqrt{3/2}$. Owing to this weighting effect, there is considerable improvement in the stochastic estimate for positive u' events through the addition of terms beyond the linear one. Because of the appreciable skewness in the probability distribution function, the imbalance between contributions at positive and negative levels of u' causes a slight degradation in estimates of negative u' events, despite the additional degrees of freedom allowed by higher-order stochastic event vector models. This effect becomes more important to the accurate estimation of coherent motions as skewness increases closer to the wall. Thus, merely increasing the order of a polynomial model of the event vector will not necessarily increase the accuracy of a least-squares stochastic estimate—estimates of different order should be used selec-

tively according to the event and estimated error in order to optimize accuracy.

B. Choice of distribution for the maximum likelihood estimation

When the deviations of data from models are described more accurately by non-Gaussian distributions, the model coefficients may be deduced by employing iterative multidimensional minimization schemes such as Powell's method to solve (3). The stochastic estimate of the time-delayed conditional average of u' is shown in Fig. 7 when the stochastic model coefficients are deduced by minimizing the sum of the absolute deviations of data from the stochastic model. This particular fit may be considered robust, in the sense that it is far less sensitive than least-squares fits to data which deviate substantially from the stochastic model. As such, it is well suited to building stochastic models of coherent motions from noisy experimental data in which outlying values may be deemed less trustworthy than those close to the estimate. This fit is optimal when the data follow a double exponential distribution about the model, though it is clear from Fig. 6 that this model distribution bears little resemblance to the measured distribution of the data about its mean. The plots of contours of relative uncertainties of 10% and 20% are shown for an event vector model comprising constant (nonzero), linear and quadratic coefficients [as in (2)]. It indicates improved performance in estimation accuracy over the least-squares estimates for these data, especially for events corresponding to negative values of u', though at an increased computational expense. In general, any distribution of data about a model may be assumed that provides a reasonable description of the joint distribution of

data and event vectors, such as the generalized analytical model proposed by Barndorff–Nielsen[14] and shown in Fig. 6. In practice, the advantages over simpler models appear to be small, and expectations of making stochastic estimation with 10%–20% relative accuracy over displacements of the order of an integral scale appear to be realistic (consistent with several of the studies referenced earlier). At displacements of this scale, spatiotemporal and event-vector expansions therefore provide comparable models of the average condition of turbulent coherent motions.

C. Generalization of the event vector expansion as a Laurent series

The approach taken by previous workers in this field[3,7] to account for departures from joint normalcy between real data and event vectors has been to extend the model through the addition of higher-order polynomial terms in $E(x,t)$, consistent with the form of a Taylor series expansion about $g_i(x,t)$ in $E(x,t)$ space. However, this approach loses some of its appeal when the stochastic model coefficients are deduced by carrying out minimizations over all measured events, regardless of whether or not their deviations from their mean are differential in scale. A more general approach, which retains much of the same simplicity, is to compensate for deviations from joint normalcy by expanding the linear model through the addition of polynomial terms, both positive and negative in power. The resultant model is then a Laurent series in $E(x,t)$ which, when truncated at low positive and negative orders, leads to modeling of $\langle g_i(x+r,t+\tau)|E(x,t)\rangle$ as

$$\langle g_i(x+r,t+\tau)|E(x,t)\rangle$$

$$= \cdots + \frac{Z_{ij}(r,\tau)}{E_j(x,t)} + A_i(r,\tau) + B_{ij}(r,\tau)E_j(x,t)$$

$$+ C_{ijk}(r,\tau)E_j(x,t)E_k(x,t) + \cdots . \qquad (10)$$

Since the inclusion of higher-order polynomial terms has the effect of lending extra weight to the contributions of rarer events to the stochastic model coefficients, the inclusion of terms of negative order biases the stochastic model coefficients toward better representation of more frequent events. Thus one may tailor the choice of terms in the stochastic model to reflect the kinds of events one wishes to model most accurately. As an example, stochastic estimates and conditional averages of $\langle u(x,t+\tau)|u(x,t)\rangle$ are shown in Fig. 8 when conditioned on the events $u'/\sigma=\pm 0.1$. For such small displacements in event space, the estimated truncation error is of the order of the residual uncertainty in the test data, necessitating direct comparisons with the conditional average for determination of the accuracy in the stochastic estimate. Least-squares stochastic estimates are plotted for event vector models, which incorporate (i) a constant coefficient (zero), a linear, and a quadratic term; and (ii) a term of order -1, a constant (nonzero) coefficient, a linear, and a quadratic term. Although the truncated Laurent series representation admits a singularity for the case of individual data and events that are identical in value to their mean,
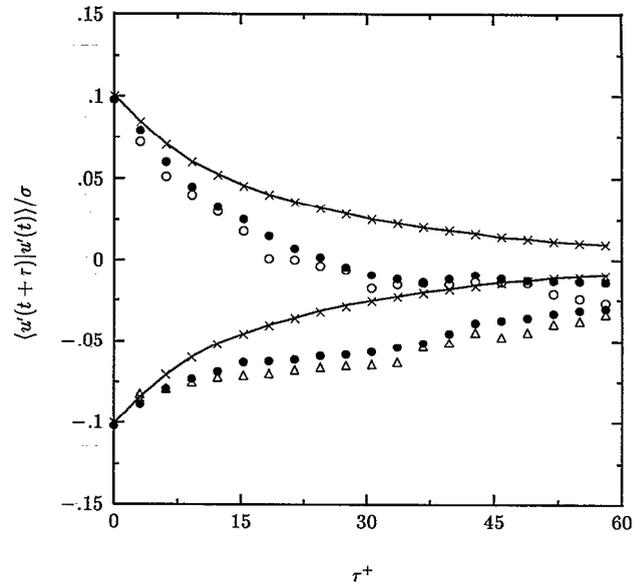

FIG. 8. Comparisons of time-delayed conditional averages and their stochastic estimates for the events $u'/\sigma=\pm 0.1$; $\times$, least-squares linear stochastic model; ———, least-squares quadratic stochastic model; ●, least-squares estimate for a Laurent series stochastic model; ○, conditional average at $u'/\sigma=0.1$; △, conditional average at $u'/\sigma=-0.1$.

none were present in the test data set. The stochastic model coefficients could therefore be evaluated by standard least-squares minimization. The close agreement between the Laurent series model and the conditional average attests to the expedience of including a term of order -1 in the stochastic model for this estimate. The disparity between the estimates of the Laurent series model and the linear and quadratic models may be attributed to the weighting of coefficients in the latter models to favor more accurate estimates of less frequent events and the realization that conditional averages of different turbulent events can depend strongly on the event under interrogation. As a consequence of this weighting, the Laurent series model yields very poor estimates for events distant from their mean. While the ability to make accurate estimates for frequent events may not be particularly important for data sets in which conditional averages of those events may be deduced from many ensembles with correspondingly low levels of uncertainty, it may be of great utility in building stochastic models to describe coherent motions conditioned on events of multiple dimension, some of which are close to their means.

D. Use of error estimates for interrogation of rare events

A further variation on standard stochastic estimation procedures is to condition the estimate on a restricted region of data space. This application is well suited to exploring whether different regions of data sets contain motions that are characterized by quite different statistics, and so might be poorly represented by a single stochastic model. For the restricted data set $g_i'(x,t)$ [a subset of

$g_i(\mathbf{x},t)$, with the corresponding event vector $\mathbf{E}'(\mathbf{x},t)]$, (2) may be recast in orthogonal polynomial form as

$$\langle g_i'(\mathbf{x}+\mathbf{r},t+\tau)\,|\,\mathbf{E}'(\mathbf{x},t)\rangle$$

$$=A_i(\mathbf{r},\tau)+B_{ij}(\mathbf{r},\tau)\,[E_j'(\mathbf{x},t)-\alpha_j(\mathbf{r},\tau)]+C_{ijk}(\mathbf{r},\tau)$$

$$\times[E_j'(\mathbf{x},t)-\beta_j(\mathbf{r},\tau)]\,[E_k'(\mathbf{x},t)-\gamma_k(\mathbf{r},\tau)]+\cdots,$$
$$(11)$$

where $\alpha_i(\mathbf{r},\tau)$, $\beta_i(\mathbf{r},\tau)$, and $\gamma_i(\mathbf{r},\tau)$ are the parameters that render the coefficients A_i, B_{ij}, and C_{ijk} independent of the choice of event vector space and so represent physical characteristics of the data. The truncation error estimate Δ_i may be deduced for this estimation model in the manner outlined in Sec. III. In a typical application, one would assess Δ_i in the stochastic estimate for an extreme event, based on all available data. This error would be compared with Δ_i' in the stochastic estimate for the same event, only based on a reduced data set, which excluded the most extreme motions. If $\Delta_i \leqslant \Delta_i'$, it would imply that the additional data were modeled as faithfully as the remainder of the data set and lent increased smoothness to the estimate, with a commensurate lowering of the error estimate. Conversely, if $\Delta_i > \Delta_i'$, one would infer that the removal of extreme data characterized by statistics that differed from the remainder of the set improved the smoothness of the estimate.

This application of stochastic estimation was tested on the turbulent boundary-layer time series for estimation of the time-delayed conditional average at $u'/\sigma = \pm 2.5$. The stochastic estimate and the relative error were computed as functions of time delay, first for the entire data set, and second, for those data for which $-2.5 < u'/\sigma < 2.5$. For the case of estimates of $\langle u(\mathbf{x},t+\tau)\,|\,u(\mathbf{x},t)\rangle$ at $u'/\sigma = 2.5$, the estimate and error deduced with each data set were indistinguishable, whereas at $u'/\sigma = -2.5$ there was appreciable reduction in the relative error when the most extreme data were removed from the set (Fig. 9). The improvement in the estimate at $u'/\sigma = -2.5$ when extreme negative u' data are removed indicates that these motions are not closely related to the more frequent events, to which the stochastic estimation model is weighted by virtue of numbers. An examination of the computed conditional averages (Fig. 10) confirms this result and reveals that events of large negative u' initially exhibit a significantly more rapid relaxation toward a long-time mean condition than other motions, which follow a more universal trend, well suited to a smooth stochastic model. The behavior of small-displacement conditional averages is also distinctly different from more extreme events, as was noted in the previous section.

The physical interpretation that can be given to such a result from a one-dimensional data set is limited, though the rapid relaxation of strongly negative u' events at $y/\delta = 0.4$ may be related to the more energetic ejection motions, consistent with their near-wall characterization in that form by Nakagawa and Nezu[16] and by Corino and Brodkey.[17] This explanation is also consistent with the findings of Bullock et al.[18] that the lowest-frequency events of the u' signal are highly correlated over large distances
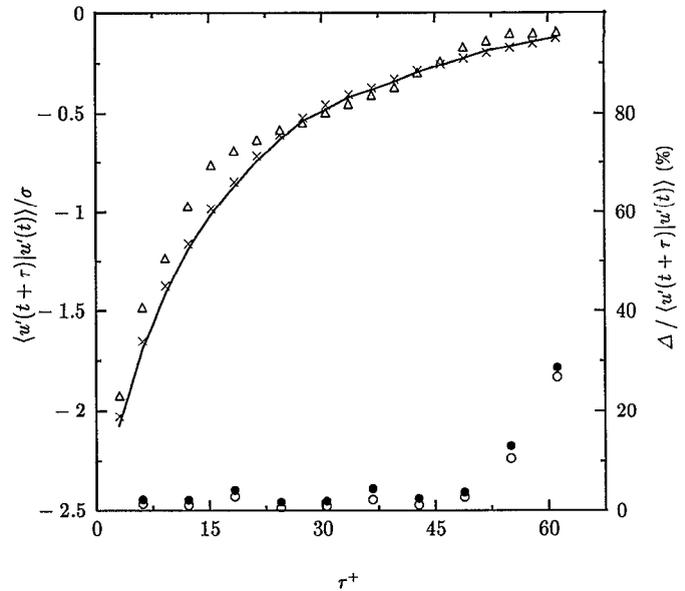

FIG. 9. Quadratic stochastic estimates and error estimates of the time-delayed conditional average of $u'(t)$ for the event $u'/\sigma = -2.5$; ———, least-squares stochastic estimate; $\times$, stochastic estimate conditioned on the restricted data set $-2.5 < u'/\sigma < 2.5$; ●, error estimate Δ for the stochastic estimate; $\bigcirc$, error estimate Δ for the conditioned estimate; $\triangle$, conditional average within a window of $\pm 0.1\sigma$.

normal to the wall. An explanation of this kind would account for the distinctly different dynamics revealed by the time-delayed conditional average of such events for durations of τ^+ of 15 following their detection, after which they appear to revert to a more universal relaxation time scale. Qualitatively similar observations of greater disparities in relaxation time scales have been reported by Adrian and Lekakis[19] in turbulent pipe flow at $Re = 50\,000$.
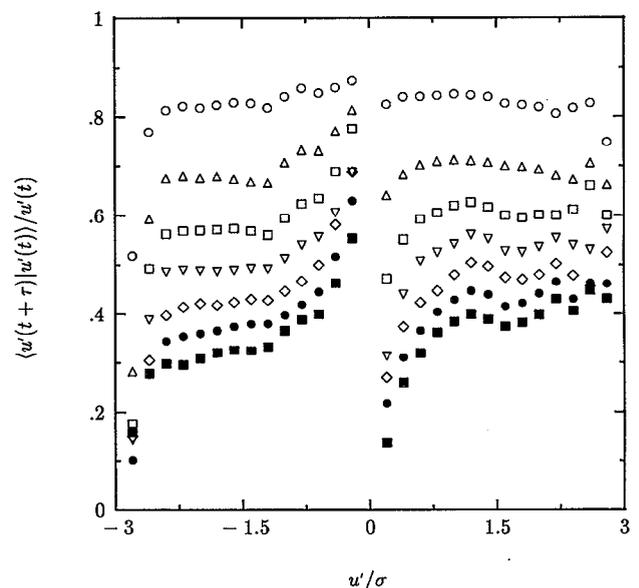

FIG. 10. Time-delayed conditional averages of $u'(t)$ for different u' events; $\bigcirc$, $\tau^+ = 3$; $\triangle$, $\tau^+ = 6$; $\square$, $\tau^+ = 9$; $\triangledown$, $\tau^+ = 12$; $\Diamond$, $\tau^+ = 15$; ●, $\tau^+ = 18$; ■, $\tau^+ = 21$.

This use of an error estimate as a guide to the smoothness of stochastic estimation models appears to be well suited to stochastic estimation of conditional averages from large multidimensional data sets, for the purposes of educing coherent motions. Once restricted regions of data space have been selected, the various estimates and errors may be computed in a single pass of the data and stored for subsequent interrogation. They may then be examined to see if the smoothed estimates of extreme events are well represented in terms of models weighted toward less extreme events, or if extreme events are distinct in character and should be examined by more selective means.

IV. CONCLUDING REMARKS

Stochastic estimation is a flexible tool that may be adapted in many ways for interrogating conditional turbulent motions. The principal contributions of this paper to its usage in fluid mechanics research lie in the extension of the technique to assess uncertainty in stochastic estimates and in its employment as a means of refining the estimation and interpretation of turbulent boundary-layer motions. In particular, it is used to isolate the relaxation of very occasional low-speed motions in the outer boundary layer, the statistical character of which would pass undetected by conventional stochastic estimation procedures. While emphasis has been placed on the stochastic estimate of the time-delayed conditional average, the extensions described in this paper may also be applied to quasidynamic reconstruction of coherent motions from databases, in which time sequences of measured event vectors associated with some characteristic feature of the flow are used to described the most likely evolution of the surrounding flow field.[5] Concurrent uncertainty estimates may then be used to assess the fidelity of the reconstructed field in describing some coherent structure.

The principal requirement that databases must satisfy to benefit from these extensions is that data vectors must be recorded at intervals in space or time that are sufficiently small (i.e., comparable in order to the Taylor microscale of length or time) to allow stochastic modeling, both in that dimension and in event vector space. This requirement should present no difficulties for well-resolved computational databases or experimentally measured time series, though it may prove more exacting for experimental measures of spatial dependence. The extensions outlined in this paper may then be used to provide accompanying estimates of accuracy in conventional applications, which employ one stochastic model for estimates of all possible events. If one is interested in making refined stochastic estimates of selected events, one may tailor the stochastic estimate to provide the best representation of those events using accuracy estimates to guide the choice of model. These refinements over existing stochastic estimation procedures allow stochastic estimation to be used as a more trustworthy tool for statistical interrogation of conditional events in turbulence databases.

[1] R. J. Adrian, "Structural information obtained from analysis using conditional vector events: A potential tool for the study of coherent structures," *Coherent Structure of Turbulent Boundary Layers*, edited by C. R. Smith and D. E. Abbott (Lehigh University, Bethlehem, PA, 1978), p. 416.

[2] R. J. Adrian, "Conditional eddies in isotropic turbulence," Phys. Fluids 22, 2065 (1979).

[3] Y. G. Guezennec, "Stochastic estimation of coherent structures in turbulent boundary layers," Phys. Fluids A 1, 1054 (1989).

[4] W. C. Choi and Y. G. Guezennec, "On the asymmetry of structures in turbulent boundary layers," Phys. Fluids A 2, 628 (1990).

[5] D. R. Cole, M. N. Glauser, and Y. G. Guezennec, "An application of the stochastic estimation to the jet mixing layer," Phys. Fluids A 4, 192 (1992).

[6] Y. A. R. Hassan, B. G. Jones, and R. J. Adrian, "Nuclear Thermal-Hydraulics Report 2," UILI-ENG, 87-5302, University of Illinois, Urbana, IL, 1987.

[7] R. J. Adrian, B. G. Jones, M. K. Chung, Y. Hassan, C. K. Nithianandan, and A. T.-C. Tung, "Approximation of turbulent conditional averages by stochastic estimation," Phys. Fluids A 1, 992 (1989).

[8] R. J. Adrian, "Multi-point optical measurements of simultaneous vectors in unsteady flow—a review," Int. J. Heat Fluid Flow 7, 127 (1986).

[9] K. Nishino and N. Kasagi, "On the quasi-coherent turbulence structures in the two-dimensional channel flow," Proceedings of the 8th Symposium on Turbulent Shear Flows, Münich, 1991, p. 28.3.1.

[10] R. J. Adrian and P. Moin, "Stochastic estimation of organized turbulence structure: Homogeneous shear flow," J. Fluid Mech. 190, 531 (1988).

[11] P. Moin, R. J. Adrian, and J. Kim, "Stochastic estimation of organized structures in turbulent channel flow," Proceedings of the 6th Symposium on Turbulent Shear Flows, Toulouse, 1987, p. 16.9.1.

[12] A. Papoulis, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, New York, 1984).

[13] R. von Mises, *Mathematical Theory of Probability and Statistics* (Academic, New York, 1964).

[14] O. Barndorf-Nielsen, "Models for non-Gaussian variation, with applications to turbulence," Proc. R. Soc. London Ser. A 368, 501 (1979).

[15] M. Farge, "Wavelet transforms and their applications to turbulence," Annu. Rev. Fluid Mech. 24, 395 (1992).

[16] H. Nakagawa and I. Nezu, "Structure of space-time correlations of bursting phenomena in an open-channel flow," J. Fluid Mech. 104, 1 (1981).

[17] E. R. Corino and R. S. Brodkey, "A visual investigation of the wall region in turbulent flow," J. Fluid Mech. 37, 1 (1969).

[18] K. J. Bullock, R. E. Cooper, and F. H. Abernathy, "Structural similarity in radial correlations and spectra of longitudinal velocity fluctuations in pipe flow," J. Fluid Mech. 88, 585 (1978).

[19] R. J. Adrian and I. C. Lekakis, "A scale gap in the structure of moderate Reynolds number turbulent pipe flow," Bull. Am. Phys. Soc. 36, 10, 2698 (1991).

2054   Phys. Fluids A, Vol. 4, No. 9, September 1992

G. J. Brereton   2054