



HAL
open science

Analysis of the filtering effect of the stochastic estimation and accuracy improvement by sensor location optimization

A. Arnault, J. Dandois, J.M. Monnier, J. Delva, J.M. Foucaut

► To cite this version:

A. Arnault, J. Dandois, J.M. Monnier, J. Delva, J.M. Foucaut. Analysis of the filtering effect of the stochastic estimation and accuracy improvement by sensor location optimization. AIAA Aviation 2016, Jun 2016, WASHINGTON, United States. 10.2514/6.2016-3257 . hal-01384220

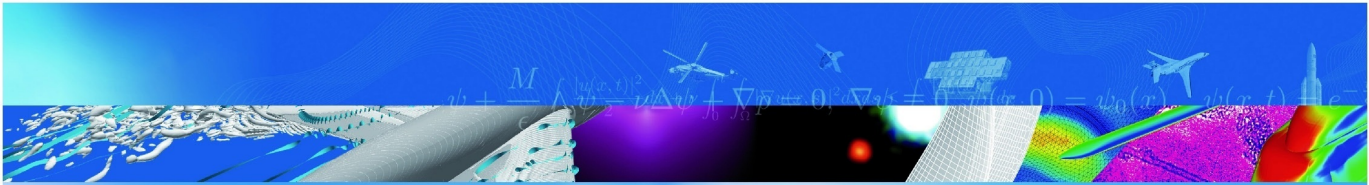
HAL Id: hal-01384220

<https://hal.science/hal-01384220>

Submitted on 19 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



COMMUNICATION A CONGRES

**Analysis of the filtering effect of the
stochastic estimation and accuracy
improvement by sensor location
optimization**

A. Arnault, J. Dandois, J.-C. Monnier,
J. Delva, J.-M. Foucaut (LML)

AIAA Aviation 2016
WASHINGTON, U.S.A
13-17 juin 2016

TP 2016-641

70 2016
ans

ONERA

THE FRENCH AEROSPACE LAB

Analysis of the filtering effect of the stochastic estimation and accuracy improvement by sensor location optimization

A. Arnault¹, J. Dandois²
Onera – The French Aerospace Lab
F-92190, Meudon, France

J.-C. Monnier², J. Delva²
Onera – The French Aerospace Lab
F-59045, Lille, France

and

J.-M. Foucaut³
LML
F-59655, Villeneuve d'Ascq, France

The reconstruction of the flow behind a backward facing step at a Reynolds number of 60,000 using Linear Stochastic Estimation (LSE) and modified is investigated. In particular, the turbulent spatial integral length scales reconstructed for several sensor configurations are studied. The reconstruction of the Proper Orthogonal Decomposition (POD) modes is also performed in order to show the limitations of the LSE reconstruction for complex flows, for which taking into account only a few POD modes is not enough to represent the flow dynamics. The importance of the sensors location on the reconstruction is also emphasized and the opportunity to use a sensor location optimization algorithm investigated.

Nomenclature

h	=	step height
d_x	=	sensors grid spacing in the longitudinal direction
d_y	=	sensors grid spacing in the vertical direction
Λ	=	turbulence spatial integral length scale
λ	=	POD mode characteristic length
R^2	=	determination coefficient
u	=	streamwise velocity
v	=	vertical velocity
e	=	$\exp(1)$

I. Introduction

The problem of accurately estimating a flow state from a few sparse measurements is currently a major challenge in various domains (aerodynamics, chemical industry...). Such estimation could clearly be used in active flow control in a closed-loop for example, or used to provide, to an operator, important information that could not be directly measured in real time. Thus a lot of efforts are directed towards the development of methods to accurately estimate a flow state from few sparse measurements.

One method, which received much attention in fluid dynamics, is the Stochastic Estimation (SE). This technique was first introduced, in the form of the Linear Stochastic Estimation (LSE), by Adrian [1] in 1977 and was used to approximate conditional averages for turbulent flow and thus identify coherent structures. Since then, SE has been

¹ PhD Student, Applied Aerodynamics Department, anthony.arnault@onera.fr.

² Research Engineer, Applied Aerodynamics Department.

³ Professor, Laboratoire de Mécanique de Lille, Ecole Centrale de Lille.

used to describe several turbulent flows such as shear layer, turbulent boundary layer, cavity flows, jets... And several extensions have been developed: Quadratic Stochastic Estimation (QSE), modified SE (SE combined with Proper Orthogonal Decomposition POD), Extended POD (EPOD), Spectral SE and Multi-Time-Delay SE (MTD-SE). In the literature on SE, the possibility of using SE as a tool to estimate a flow state has been truly investigated only recently. Taylor and Glauser [2] are the first to really identify such an application of SE. Now, most of recent works (Durguesh et al. [3], Tu et al. [4], Clark et al. [5], Hosseini et al. [6]) focus on the use of MTD-LSE, which exhibited a better accuracy than single time LSE. Recently, Lasagna et al. [7] also investigated the development of a non linear extension of MTD-SE.

However most of the literature on SE is mainly limited to applications on flows at low Reynolds number for which a few POD modes are able to capture a very large part of the turbulent kinetic energy (TKE). For instance, Tu et al. [4] applied SE to the near wake behind a flat plate model at a Reynolds number of 3,600 based on the plate thickness. In their case, almost 80% of the TKE is captured by the first two POD modes. Hudy et al. [8] used modified LSE to investigate the evolution of coherent structures in the flow downstream a backward facing step at a Reynolds number of 8,081. Nguyen et al. [9] tested several methods to estimate the flow downstream a backward facing step at a Reynolds number of 2,800 based on the step height. As concluded by Clark et al. [5], SE is expected to perform well in such conditions, but for more complicated flows, at high Reynolds number with a larger range of turbulent scales, the accuracy of the SE has yet to be better studied.

The reconstruction of the backward facing step velocity field, for a Reynolds number of 60,000 based on the step height, from wall pressure measurements was studied by Arnault et al. in ref. [10]. The velocity field estimations obtained were of low accuracy and little improvements were found in using QSE and MTD-SE. The objective of the present work is then to better characterize SE limitations. In particular, we focus ourselves on the study of the conservation of the turbulence spatial integral length scales by LSE. The impact of the type of conditional events, as well as their positioning, is also investigated.

This paper is organized as follows. Section II briefly summarizes the principles of POD and SE techniques. The experimental setup is described in section III. The study of the conservation of the integral length scales is presented in section IV and the impact of the sensor locations is discussed in section V. A quick roadmap for the use of SE is then proposed (section VI). At last, conclusions are given in section VII.

II. Mathematical background

A. Proper Orthogonal Decomposition (POD)

POD or Karhunen-Loeve decomposition is a flow decomposition technique proposed by Lumley [11] to identify coherent structures. The idea is to consider that these structures have the largest projection on the velocity field (according to the mean square method). The POD basis is then the collection of spatial vectors $\Phi(\mathbf{x})$ that are the eigenmodes of the equation:

$$\int R_{uu}\Phi_i(\mathbf{x})d\mathbf{x} = \sigma_i \Phi_i(\mathbf{x}) \quad (1)$$

where R_{uu} is the two-point spatial correlation matrix for the time-dependent velocity vector field $\mathbf{u}(\mathbf{x},t)$ and σ_i is the eigenvalue associated with the POD mode i .

Velocity vector field can then be decomposed on the POD basis:

$$\mathbf{u}(\mathbf{x},t) = \sum_{i=1}^{N_{\text{POD}}} a_i(t) \Phi_i(\mathbf{x}) \quad (2)$$

Where N_{POD} is the total number of POD modes and $a_i(t)$ the POD coefficients that can be expressed as:

$$a_i(t) = (\mathbf{u}(\mathbf{x},t), \Phi_i(\mathbf{x})) \quad (3)$$

where $(. , .)$ represents the inner product.

In the present work, the snapshot method [12] is used to compute the POD modes and coefficients.

B. Stochastic Estimation (SE)

SE principle is to consider that the estimation of a quantity, for instance a scalar velocity field $u(\mathbf{x},t)$, under an event \mathbf{E} (called the conditional event) is similar to a conditional average:

$$\tilde{u}(\mathbf{x},t) = \langle u(\mathbf{x},t) | \mathbf{E} \rangle \quad (4)$$

If the quantity u is assumed to be continuous and zero mean, then its Taylor expansion around the value of \mathbf{E} is:

$$\tilde{u}(\mathbf{x}, t) = \sum_{i=1}^N A_x(y_i) \mathbf{E}(y_i, t) + \sum_{i=1}^N \sum_{j=1}^N B_x(y_i, y_j) \mathbf{E}(y_i, t) \mathbf{E}(y_j, t) + \dots \quad (5)$$

N is the number of y_i points where \mathbf{E} is known (for instance the number of sensors used).

Keeping only the first order terms in the Taylor expansion leads to LSE. Keeping first and second order terms leads to QSE. In both cases, A_x and B_x coefficients are calculated through the minimization of the mean square error the SE and the true value on a training set of data:

$$e(\mathbf{x}) = \langle (\tilde{u}(\mathbf{x}, t) - u(\mathbf{x}, t))^2 \rangle \quad (6)$$

Once the coefficients calculated on the training data set, it is possible to estimate the quantity u even outside the training data set without any knowledge of u . Thus the estimation of u **outside of the training set is referred as the prediction of u** , whereas the estimation of u **in the training set is referred as the reconstruction of u** .

From the way LSE is introduced here, one can clearly see its relationship to the Ordinary Least Square regression [5]. For more details, one can refer to [5] or [13].

Modified LSE or LSE-POD consists in the estimation of the POD coefficients instead of the entire velocity field. The multi-time-delay technique consists in using several realizations of the conditional event \mathbf{E} , not only in space, but also in time. The estimation of a scalar velocity field $u(\mathbf{x}, t)$ then becomes:

$$\tilde{u}(\mathbf{x}, t) = \sum_{i=1}^N \sum_{j=1}^{N_d} A_x(y_i, \tau_j) \mathbf{E}(y_i, t - \tau_j) \quad (8)$$

where N_d is the number of delays used. More details on MTD-SE can be found in [3].

To compare the accuracy of several estimations, the determination coefficient R^2 is used and defined by:

$$R^2 = 1 - \frac{\int_D \langle \sum_{i=1}^{n_c} (\tilde{u}_i(\mathbf{x}, t) - u_i(\mathbf{x}, t))^2 \rangle d\mathbf{x}}{\int_D \langle \sum_{i=1}^{n_c} u_i(\mathbf{x}, t)^2 \rangle d\mathbf{x}} \quad (10)$$

where n_c is the number of velocity components estimated and D the spatial domain. In the case where POD is used, we also define the determination coefficient of the POD mode i by:

$$R_{\text{POD Single, } i}^2 = 1 - \frac{\langle (\tilde{a}_i(t) - a_i(t))^2 \rangle}{\langle a_i(t)^2 \rangle} \quad (11)$$

III. Experimental setup

The experiment was conducted in the $0.3 \text{ m} \times 0.3 \text{ m} \times 2.64 \text{ m}$ boundary layer wind tunnel of the Onera Lille center. This wind tunnel is of Eiffel type with a variable frequency driven motor capable of producing free-stream velocities of 10 to 40 m/s, for a turbulent intensity lower than 2%. The backward facing step height is $h = 30 \text{ mm}$. The Reynolds number based on h is approximately 60,000 and the Mach number about 0.09. The measured boundary layer thickness at $0.3 h$ upstream of the step is approximately $0.38 h$.

Particle Image Velocimetry (PIV) and pressure transducers (Kulite XCQ-80-5PSID) measured the separated flow field downstream of the backward facing step and the unsteady pressure at the wall, respectively. The PIV experiment used a twin 120 mJ Nd:YAG lasers and two CCD cameras (in order to cover the full separated flow from the step to the reattachment point). All PIV recordings have been processed by an in-house software. The evaluation of these images was performed with a cross-correlation scheme using standard FFT with multi-pass, interrogation window shift and a final window size of 32×32 pixels with 50% overlap, corresponding to a resolution of about 1.2 mm in both directions.

Three sets of 3400 uncorrelated PIV images have been acquired along with simultaneously sampled pressure signals. PIV images are sampled at 5 Hz (spanning about 20 vortex shedding periods) and the pressure signals are sampled at 20 kHz. The acquired data are split into a training set consisting of the first two sets of acquisition and the last one is used as a validation set. Thus, the training set contains 6794 PIV fields (due to synchronization considerations, in order to be able to use multi-time-delay methods, the first three PIV fields of each acquisition set are disregarded), and the corresponding pressure measurements from the 17 Kulites. The validation set contains

3197 PIV fields and the corresponding pressure measurements. In the following, all the results are obtained on the validation set.

In addition to the previous campaign, high speed PIV has also been conducted. A twin pulsed 20 mJ Nd-YLF lasers was used and PIV snapshots were recorded by two CMOS cameras. The resolution of the PIV snapshots, after post-processing, is approximately 2.3 mm. The PIV acquisition frequency was set at 2 kHz. Four sets of 6144 PIV images have been acquired. A vortex shedding period covers about 20 snapshots. As previously, data are split between a training set and a validation set. The training set contains 18432 snapshots and the validation set 6144.

IV. Turbulence Spatial Integral Length Scales

The goal of this first part is to study the conservation of the turbulence spatial integral length scales by the LSE. More particularly, the impact of the conditional events used on the conserved turbulence spatial length scales is investigated.

A. Backward facing step turbulence spatial integral length scales

The turbulence spatial integral length scales of one velocity component are calculated from the spatial autocorrelation function according to equation [14]:

$$\Lambda = \int_0^{\infty} R_{ii}(r) dr \quad (12)$$

where the double- i subscript indicates the autocorrelation function of the component i defined by:

$$R_{ii}(r) = \frac{\langle u_i(x_i, t) u_i(x_i + r, t) \rangle}{\sqrt{\langle u_i^2(x_i) \rangle \langle u_i^2(x_i + r) \rangle}} \quad (13)$$

and r is the distance between two points in the flow. If the direction of r is parallel to u_i the autocorrelation function is called longitudinal, if not, it is called transverse. Following O'Neill et al. [14], the integration domain of eq. (12) was chosen to range from 0 to the point where the autocorrelation goes under $1/e$.

Maps of the longitudinal and transverse integral length scales for u are displayed in Fig. 1 and for v in Fig. 2. The discontinuities that can be observed around $x/h = 3.5$ comes from overlapping flaws between the two PIV fields that form the image. The largest integral length scales (about one half the step height) are mainly located above the recirculation and downstream of the reattachment point (located at $x/h = 5.6$). They result from the vortex shedding downstream of the step and not from the freestream flow since the values upstream of the step are close to zero. This indicates that large coherent structures are formed downstream of the step and located in this area.

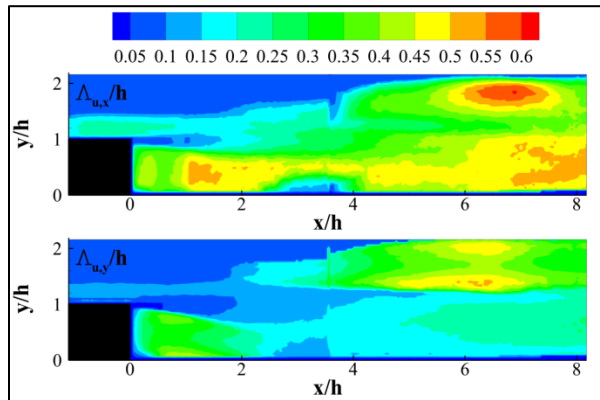


Fig. 1: Longitudinal (top) and transverse (bottom) turbulence spatial integral length scale for the streamwise velocity u

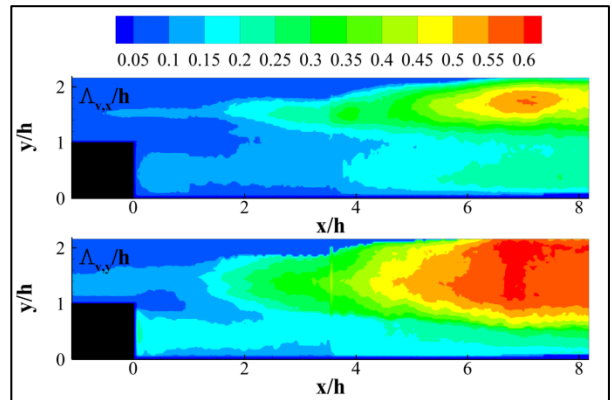


Fig. 2 : Transverse (top) and longitudinal (bottom) turbulence spatial integral length scale for the vertical velocity v

B. LSE using wall pressure measurements

Following the work of Hudy [8], who investigated the coherent structures in a backward facing step flow, at $Re_h = 8081$, using LSE and wall pressure measurements from microphones, the first situation tested uses 17 wall pressure measurements as conditional events for the LSE. As already explained, in our case, the Reynolds number is higher ($\sim 60,000$), thus the flow is expected to present more energy at high frequencies and smaller turbulence spatial integral length scales than the one studied by Hudy.

As stated in the introduction, the estimation accuracy was found to be poor and the use of QSE, LSE-POD, QSE-POD or multi-time delays SE did not lead to any drastic improvement, with determination coefficient R^2 ranging from 7.9% to 9.3% in the best case using multi-time-delays LSE (with Tikhonov regularization, see [10] for more details). Fig. 3 and Fig. 4 show the longitudinal and transverse turbulence spatial length scale for u and v predicted by LSE (validation set) (black triangles indicate pressure sensors locations). Generally speaking, the integral length scales of the estimated fields are larger than the original ones everywhere in the flow. Patterns present in the original maps in Fig. 1 and Fig. 2 cannot be identified in the maps obtained from the estimated fields. The turbulence spatial integral length scales are not conserved by this LSE.

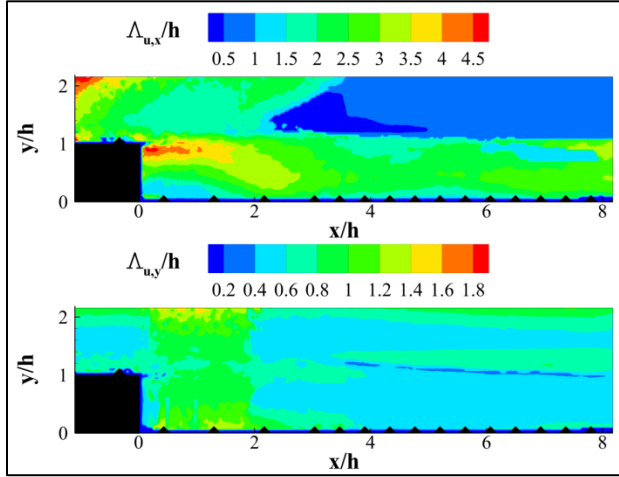


Fig. 3: Longitudinal (top) and transverse (bottom) turbulence spatial integral length scale for the streamwise velocity u estimated by LSE (17 wall pressure sensors)

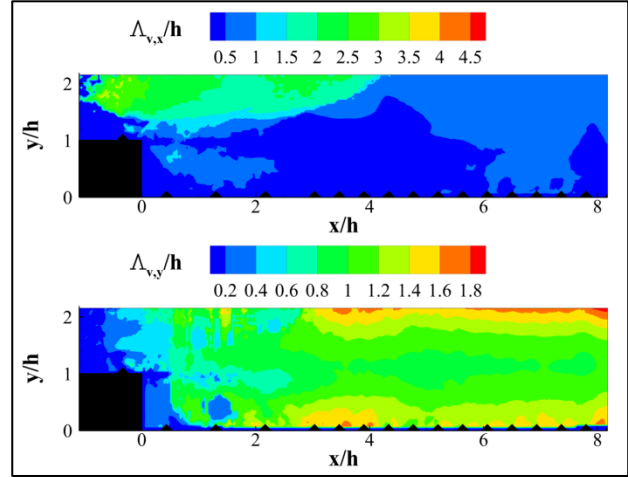


Fig. 4: Transverse (top) and longitudinal (bottom) turbulence spatial integral length scale for the vertical velocity v estimated by LSE (17 wall pressure sensors)

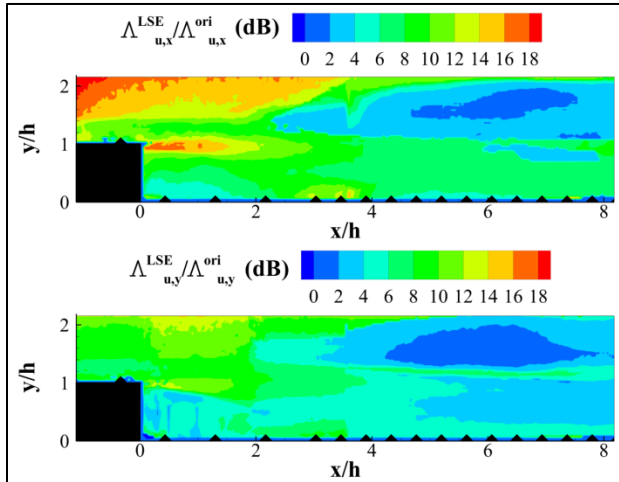


Fig. 5: LSE prediction to original turbulence spatial integral length scale for u ratio (top: longitudinal, bottom: transverse)

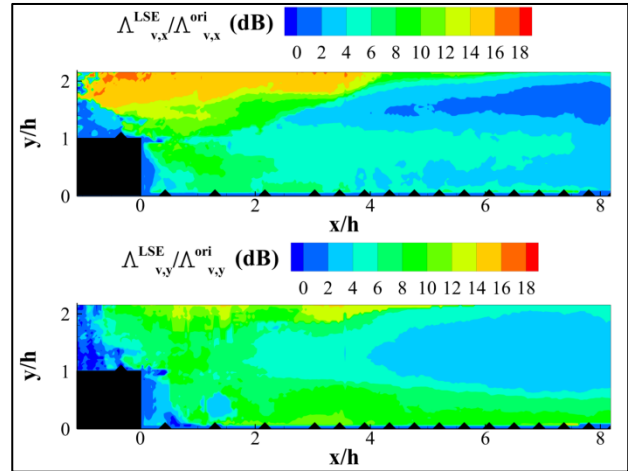


Fig. 6 : LSE prediction to original turbulence spatial integral length scale for v ratio (top: transverse, bottom: longitudinal)

In order to better compare the integral length scales obtained from the predicted fields with the original one, their ratio is displayed in Fig. 5 and Fig. 6 and expressed in decibel (dB). In most regions (recirculation area, downstream of the reattachment point and above the recirculation downstream $x/h = 3$), the length scales have been increased by a factor of 2 (3 dB) to 6 (7.8 dB). In the four maps, the area of lowest increase of the integral length scales, is located above the recirculation, for $x/h \geq 4$, and corresponds to the region of highest integral length scale in the original data. This area is the one where the integral length scales are the best conserved and actually corresponds to the region where the LSE performs the best (see Arnault et al. [10]). In this configuration, using wall

pressure measurements, the LSE is only able to estimate the largest integral length scales and the smallest ones are filtered. The QSE was also investigated. It leads to a marginally better estimation of the length scales. However, it is not able to decrease the order of magnitude of the shortest length scales estimated. Therefore, in the following, we focus on the study of the LSE case.

C. LSE using velocity field sensors

Since the use of wall pressure sensors leads to a poor conservation of the turbulence spatial integral length scales, it was decided to investigate the characteristics of the LSE using velocity sensors in the flow. To do so, eight grids of streamwise velocity sensors (extracted from the PIV measurements) are employed. Table 1 summarizes the characteristics of each grid. The initial point (the most upstream and close to the wall) location is given as well as the spacings in the streamwise and vertical directions (see Fig. 7 and 8 for sensors locations). Table 2 compares the determination coefficient of each configuration. It can be seen that determination coefficients up to 62% can be reached with the finest grid. With the coarsest grids, which contains 21 sensors (only 4 more sensors than there were pressure sensors in the previous case), a determination coefficient of 21% is reached, which is more than two times the one obtained with wall pressure measurements.

Number of sensors	Initial point ($x/h, y/h$)	Streamwise spacing (d_x/h)	Vertical spacing (d_y/h)
462	(0.19, 0.15)	0.19	0.19
240	(0.19, 0.19)	0.266	0.266
126	(0.19, 0.15)	0.38	0.38
56	(0.19, 0.15)	0.57	0.57
33 (Top)	(0.19, 0.53)	0.76	0.76
33 (Bottom)	(0.19, 0.15)	0.76	0.76
21 (Top)	(0.19, 0.53)	1.14	0.76
21 (Bottom)	(0.19, 0.15)	1.14	0.76

Table 1: Sensor grid characteristics

Number of sensors	Validation set R^2	Shortest $\Lambda_{u,x}$		Shortest $\Lambda_{u,y}$		Shortest $\Lambda_{v,x}$		Shortest $\Lambda_{v,y}$	
		$\Lambda_{u,x}/h$	$\Lambda_{u,x}/d_x$	$\Lambda_{u,y}/h$	$\Lambda_{u,y}/d_y$	$\Lambda_{v,x}/h$	$\Lambda_{v,x}/d_x$	$\Lambda_{v,y}/h$	$\Lambda_{v,y}/d_y$
462	62%	0.16	0.88	0.17	0.88	0.15	0.79	0.17	0.88
240	56%	0.2	0.75	0.19	0.73	0.18	0.69	0.2	0.75
126	47%	0.27	0.71	0.23	0.61	0.23	0.61	0.22	0.57
56	35%	0.31	0.55	0.28	0.5	0.27	0.47	0.3	0.53
33 (Top)	27%	0.36	0.47	0.33	0.44	0.28	0.37	0.35	0.46
33 (Bottom)	28%	0.35	0.46	0.35	0.46	0.3	0.4	0.35	0.46
21 (Top)	20%	0.4	0.35	0.35	0.46	0.3	0.26	0.37	0.48
21 (Bottom)	21%	0.37	0.32	0.33	0.44	0.32	0.28	0.37	0.48

Table 2 : Determination coefficient and turbulence spatial integral length scales, for the validation set, of several LSE using different sensor grids

For brevity, only figures of the longitudinal integral length scale of u are displayed for each grid. Fig. 7 and Fig. 8 show the maps of this integral length scale for the eight different grids of sensors used for the LSE. From these figures, it can be observed that the coarsening of the sensors grid leads to a deterioration of the integral length scales conservation. Especially, the shortest integral length scale increases with the sensors spacing. Even with the finest grid, the integral length scales are not perfectly conserved and are increased in the entire domain. Therefore they are

also increased in the area of original largest integral length scales. These observations are also valid for the three other integral length scales $\Lambda_{u,y}$, $\Lambda_{v,x}$, $\Lambda_{v,y}$ (not shown here for brevity)

The approximate orders of magnitude of the shortest integral length scales estimated (found downstream of the step) are summarized in Table 2, for each grid. To determine these values, data at the sensor locations are disregarded, since the estimation at this point is perfect. The data are normalized by the step height h or by the sensors grid spacing, d_x or d_y , in the direction of the integral length scale calculation. As previously observed, coarsening the grid leads to an increment of the shortest integral length scale predicted. An interesting fact appears when looking at the ratio between the predicted integral length scale and the sensor grid spacing. This ratio ranges from 0.28 to 0.88 and increases when the sensor grid is refined. Thus, the shortest integral length scale predicted by the LSE is not proportional to the distance between the sensors. Large integral length scales can be recovered by LSE with sensor spacing larger than the scales. But the smallest integral length scales seems to require sensor spacing of the order of magnitude of the turbulent scales.

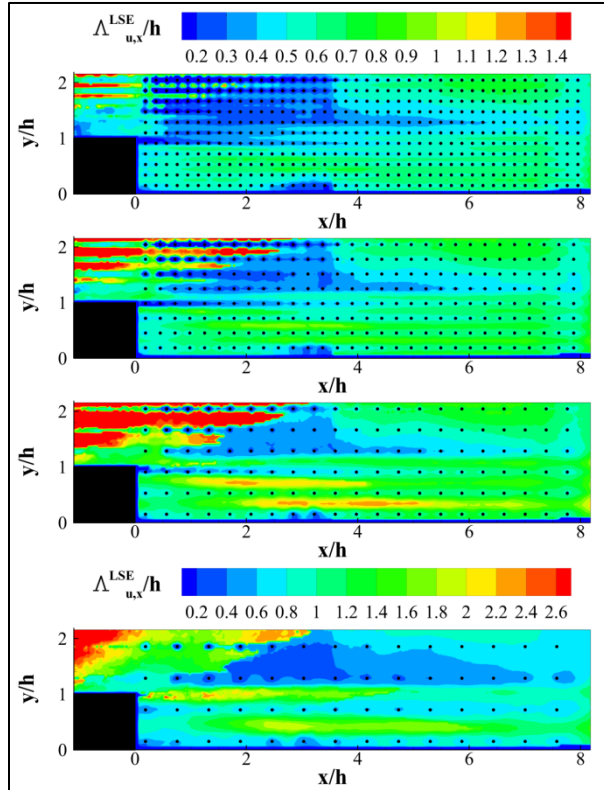


Fig. 7 : Longitudinal turbulence spatial integral length scale for u predicted by LSE for grids of 462, 240, 126 and 56 sensors (from top to bottom)

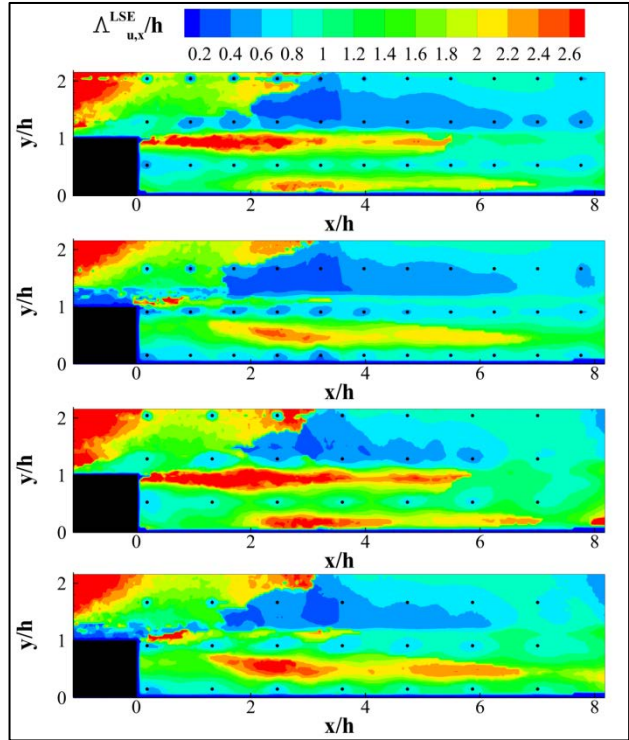


Fig. 8 : Longitudinal turbulence spatial integral length scale for u predicted by LSE for two grids of 33 sensors and two grids of 21 sensors (from top to bottom)

D. Study through POD analysis

To complete the previous study, the estimation of the POD modes of the flow field through LSE is investigated. In the following the modified LSE (or LSE-POD), as defined in section II, is used. The POD is computed using the fluctuating velocity fields from the training set (only the streamwise and vertical velocity components are available). Fig. 9 shows the energy spectrum of the decomposition for the first one hundred modes, containing approximately 70% of the TKE. This spectrum shows that an important part of the TKE is contained in high order modes. To obtain 90% of the TKE, more than 650 POD modes are required which explains why it is difficult to accurately estimate the velocity field through LSE in the present case.

The determination coefficient $R^2_{\text{POD Single}}$ of the first one thousand POD modes is plotted in Fig. 10 for the grids of 462, 240, 126 and 56 sensors. $R^2_{\text{POD Single}}$ for the first one hundred POD modes is also plotted in Fig. 11 for the grids containing 33 and 21 sensors. In the present case, the general trend is that the higher the POD mode is, the lower the determination coefficient is. In addition, for one POD mode, the coarser the sensor grid is, the lower the

determination coefficient is. Therefore, it is more difficult to estimate a high order POD mode with LSE than a low order one. Furthermore high order POD modes require more sensors to be correctly estimated. As such, only the grid with 462, 240 and 126 sensors are able to estimate some POD modes with a determination coefficient higher than 80%.

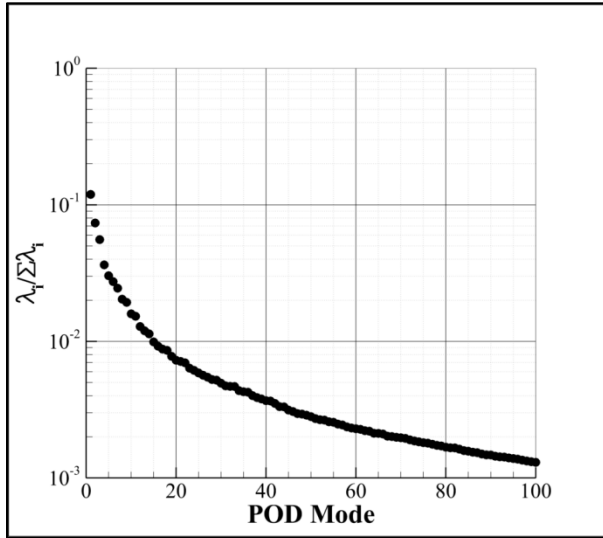


Fig. 9: Energy spectrum of the first hundred POD modes

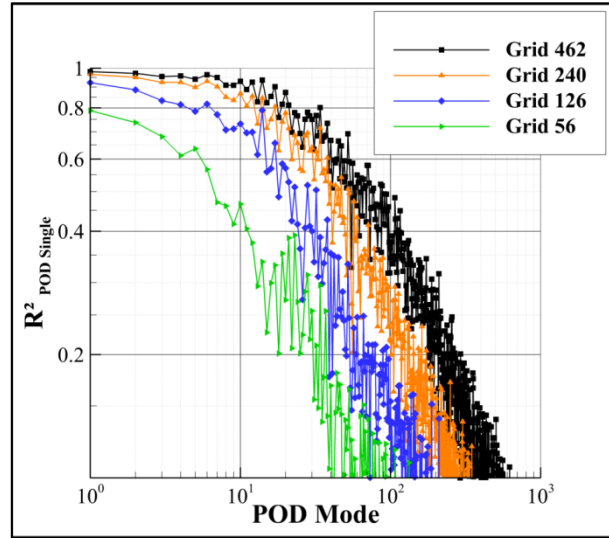


Fig. 10 : Determination coefficient as function of the POD mode predicted (estimation of validation set data)

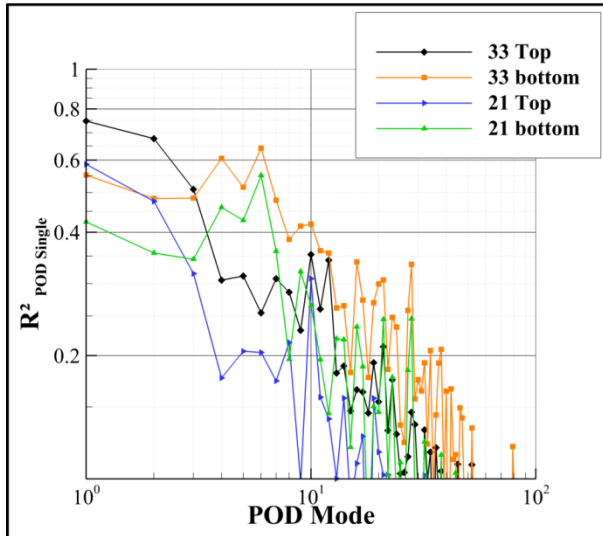


Fig. 11 : Determination coefficient as function of the POD mode predicted

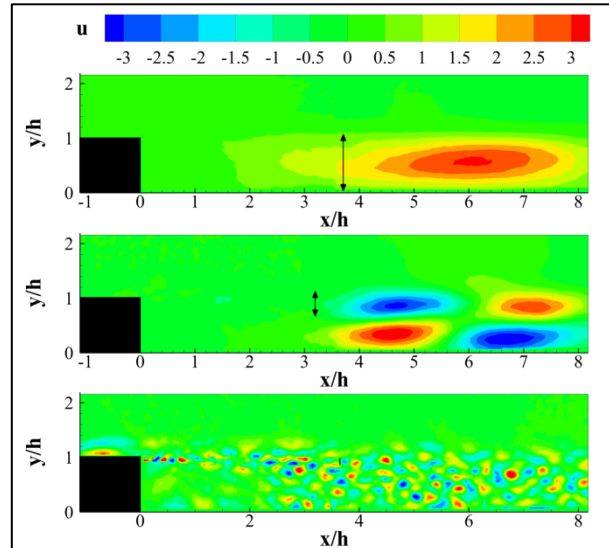


Fig. 12 : Streamwise component of the velocity POD modes 1, 7 and 400 (from top to bottom). Black arrows illustrate the characteristic length of the mode

To study the link between the sensor grid used and the predicted length scales, a characteristic length λ is associated to each POD mode. This length λ is the length for which 95% of the TKE of the POD mode is conserved after having been spatially filtered by an ideal low-pass filter of cut-off frequency $1/\lambda$ in the spatial Fourier domain. Mathematically, the characteristic length λ of each POD mode i is defined by:

$$\arg \max_{\lambda \in \mathbb{R}^+} \left(\frac{|\iint H(\zeta, \xi) \widehat{\Phi}_i(\zeta, \xi) d\zeta d\xi|^2}{|\iint \widehat{\Phi}_i(\zeta, \xi) d\zeta d\xi|^2} \geq 0.95 \right) \quad (14)$$

where $\widehat{\Phi}_i(\zeta, \xi)$ is the spatial Fourier transform of the POD mode i and $H(\zeta, \xi)$ is the ideal low-pass filter transfer function defined by:

$$H(\zeta, \xi) = \begin{cases} 1 & \text{if } D(\zeta, \xi) < 1/\lambda \\ 0 & \text{if } D(\zeta, \xi) \geq 1/\lambda \end{cases} \quad (15)$$

with $D(\zeta, \xi)$ the distance norm in the Fourier domain.

This length can be interpreted as being the shortest length of the spatial structures contained in the POD mode. Structures of smaller length count for less than 5% of the POD mode TKE. It is not a turbulent integral length scale. In the present case, this length generally decreases when the POD mode rank increases. Fig. 12 shows the streamwise velocity of the POD mode 1, 7 and 400, for which the characteristic lengths are respectively $1.1 h$, $0.56 h$ and $0.13 h$. For the 1st POD mode and the 7th, the characteristic length corresponds well to the approximate height of the structures, but is clearly smaller than their length. For the POD mode 400, the turbulent structures are more isotropic and have, globally, the same height and length which matches λ . Therefore, λ seems to correctly approximate the shortest length contained in the POD mode.

If, for example, we arbitrarily choose a threshold of 80% on the determination coefficient of one POD mode, above which it is considered to be satisfactorily predicted, then the grid 462 is able to correctly predict the POD modes up to the mode 21, the grid 240 up to 14 and the grid 126 up to 6 (the other grids do not satisfy this condition). Modes 21, 14 and 6 are associated with the characteristic lengths $0.49 h$ ($2.6 d$), $0.55 h$ ($2 d$) and $0.73 h$ ($1.9 d$) (where $d = d_x = d_y$) respectively. For each grid, one has to expect not being able to predict, with high accuracy structures shorter than these characteristic lengths. The same effect, as with the turbulence spatial length scales, is seen here where the characteristic length, of the last POD mode predicted with a R^2 higher than 80%, is not proportional to the grid spacing. When the grid is already quite dense, refining it leads to fewer improvements than the refinement of a coarse grid.

Number of sensors	Cut-off rank (10% threshold on R^2)	Associated characteristic length λ	
462	300	$0.14h$	$0.74d$
240	225	$0.17h$	$0.64d$
126	125	$0.24h$	$0.63d$
56	50	$0.32h$	$0.56d$
33 (Top)	30	$0.44h$	$0.58d$
33 (Bottom)	35	$0.37h$	$0.49d$
21 (Top)	14	$0.55h$	$0.48d_x$
21 (Bottom)	21	$0.49h$	$0.43d_x$

Table 3 : Cut-off rank of the predicted POD modes and characteristic length associated for the several grids of sensors used.

Similarly, it is possible to choose a threshold on POD modes R^2 , under which the POD mode can no longer be considered correctly predicted. We choose here a 10% threshold defining, for the 8 grids, a cut-off rank for the POD modes which are summarized in Table 3. The coarsest grids are unable to predict POD modes with characteristic length smaller than half the step height. And the finest grid is only able to predict POD modes with characteristic length higher than a tenth of the step height. It can be observed, when the characteristic length of the cut-off POD mode is expressed in sensors spacing d , that the characteristic length decreases when the grid is coarsening.

From Fig. 10 and Fig. 11, it is seen that the determination coefficient rapidly decreases with the POD mode rank. For the grid 462 to 126, the accuracy of the prediction of a POD mode quickly deteriorates for POD mode of rank higher than 30. For the other grids, the deterioration starts even before the rank 10. The SE, in this situation, seems to be able to predict correctly only the first POD modes that contain the largest scales of the flow. This fact is also seen for the estimation using the wall pressure measurements with LSE-POD (see Fig. 17). But LSE-POD performs

badly for any POD modes, with a R^2 lower than 25%. The cut-off rank is about 10 and the 10th POD mode has a characteristic length of about $0.6h$.

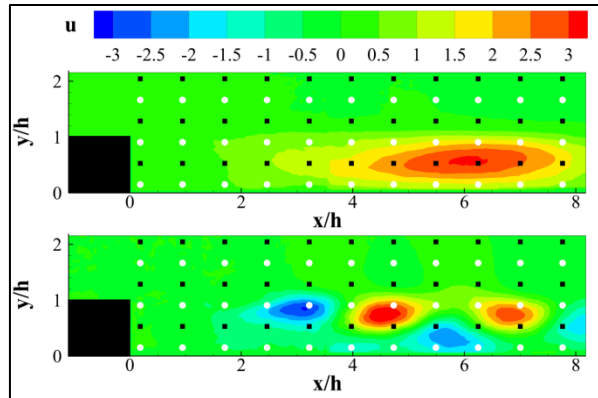


Fig. 13: Streamwise component of the velocity POD modes 1 and 6. The black squares correspond to the grid 33 “top” sensors and the white circles to the grid 33 “bottom”

Looking more closely to Fig. 11, one can observe some discrepancies of the prediction accuracy of the first 10 POD modes for the two grids made of 33 sensors and the two grids made of 21 sensors. According to the location of the grid, POD modes are predicted with different level of R^2 . Grids with sensors located at half of the step height (33 and 21 “top”), up to the top boundary of the estimated domain, perform better on the first two POD modes than grids with sensors located near the wall (33 and 21 “bottom”). On the contrary, the grids 33 and 21 “bottom” seem to better predict POD modes of rank higher than 4. The reasons for this behavior are investigated in the following part.

V. Sensor locations impact

Differences in the estimation accuracy of the POD modes between two grids, with the same number of sensors and the same spacing, confirm that the sensor grid spacing is not the only parameter impacting on the length scales estimated by LSE. In addition, the non-proportionality of the smallest turbulence spatial length scale estimated with the grid spacing also indicates that refining a uniform grid of sensors may not be the best way to capture smaller and smaller length scale structures of the flow through SE.

In 2004, Cohen et al. [15] proposed a heuristic method to place the sensors in order to estimate POD coefficient using LSE. Their method uses a POD of the possible sensors and they advocated locating the sensors at the maxima and minima of the sensors POD modes. In our case, when using velocity measurements from the PIV data as sensors, they should be located at the maxima and minima of the velocity POD modes.

The first and sixth modes are displayed in Fig. 13 where the sensors of grids 33 “top” and “bottom” are also plotted. This figure shows that sensors of the grid 33 “top” are located close to the maxima of the first POD mode for the streamwise velocity (it is also true for the vertical velocity). On the contrary, the grid 33 “bottom” has sensors above and under the maxima of the first mode and they are clearly farther to the maxima than the ones of the grid 33 “top”. It is then coherent to observe in Fig. 11 that the first POD mode is better predicted by the grid 33 “top” than by the grid 33 “bottom”. The opposite situation happens for the POD mode 6, which is better predicted when using the grid 33 “bottom”. If the two grids have around the same number of sensors close to extrema of the POD mode 6 for the vertical velocity (not shown here for brevity), the grid 33 “bottom” has sensors closer to the extrema of the POD mode 6 streamwise velocity, than the grid 33 “top”. Extrema of the POD modes seems to be indeed a better location for the sensors. And then, to correctly estimate a POD mode, it appears to be important to dispose of sensors near the extrema of the mode.

As Fig. 12 and Fig. 13 show, the higher the rank of a POD mode is, the more it possesses extrema. Therefore, one can expect that the number of sensors necessary to estimate a POD mode, with a certain level of R^2 , increases with the POD mode rank. Table 4 compares the determination coefficient obtained for POD mode 1, 2, 5 and 10, when using LSE with one sensor located at the extrema of the POD mode. It clearly appears, as expected, that the determination coefficient decreases with the POD mode rank and that more sensors are necessary to estimate higher POD modes at the same level of R^2 . This confirms the difficulty that arises when trying to estimate small scale structures from a limited number of measurements through SE. In addition, the situation is even more unfavorable

when the sensors, for example unsteady pressure transducers, can only be placed at the wall and thus not close to the extrema of the POD modes to estimate.

POD mode	R^2 using only one sensor	Number of sensors to reach 50% of R^2
1	37%	2
2	30%	2
5	20%	4
10	17%	7

Table 4: Determination coefficient of the reconstruction of several POD modes when using only one streamwise velocity sensors located at the extremum of the POD mode and number of streamwise velocity sensors required to reach 50% of R^2

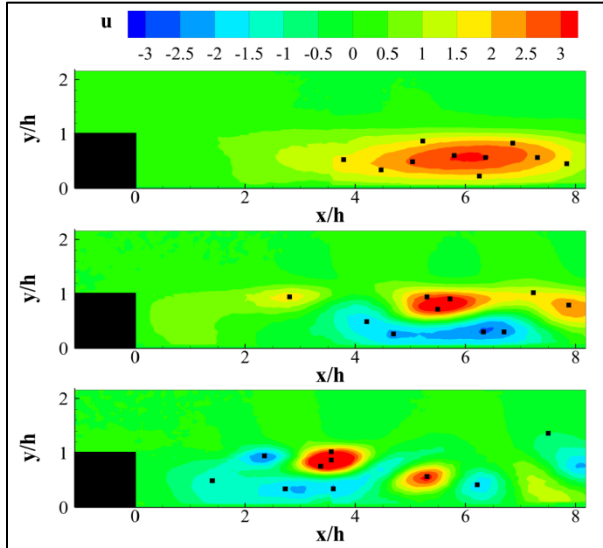


Fig. 14: Sensor locations obtained by the optimization algorithm for mode 1, 5 and 10 separately (from top to bottom). The streamwise velocity of the modes is displayed.

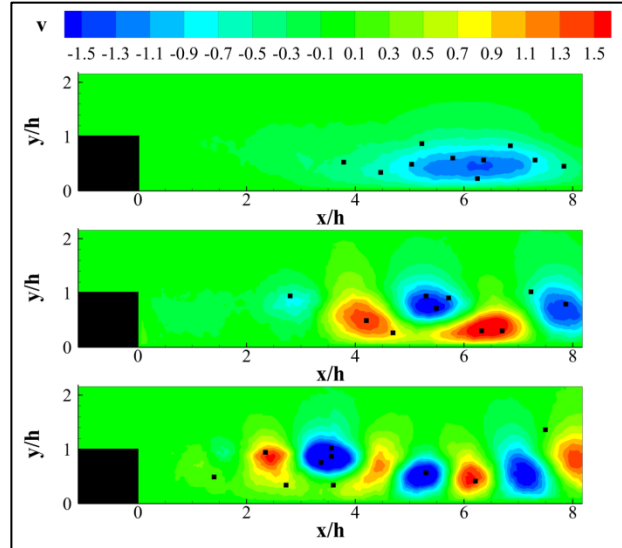


Fig. 15: Sensor locations obtained by the optimization algorithm for mode 1, 5 and 10 separately (from top to bottom). The vertical velocity of the modes is displayed.

If Cohen et al. proposed to use a cost function to choose sensors in order to estimate an ensemble of several POD modes, it was decided to investigate the use of a sensor location optimization algorithm introduced by Muradore et al. [16]. The algorithm is a data driven one, based on partial least square regression. The main idea behind the algorithm is that, if data from several sensors are available, one can choose among them such as to pick the one with the highest correlation with the data to estimate (that has not yet been explained by the previously chosen sensors). Details on the algorithm can be found in reference [16]. In our case the possible sensors to choose are the streamwise velocity data for the entire PIV domain and the data to estimate is one or several POD modes. This algorithm can also be used to determine the number of sensors necessary to reach a certain value of $R^2_{\text{POD Single}}$. The numbers of sensors necessary to reach 50% of $R^2_{\text{POD Single}}$ for several POD modes are reported in Table 4. The results confirm that more and more sensors are necessary to obtain a good estimation as its rank increases.

To verify if the sensor locations given by the algorithm follow Cohen's conclusion, the optimal locations for the reconstruction of each POD mode are computed. Fig. 14 and Fig. 15 show the locations of the first ten sensors chosen by the algorithm for the POD modes 1, 5 and 10 separately. These locations match quite well the extrema areas of each POD mode, thus confirming that they are the areas of highest correlation between the streamwise velocity data and the POD mode. The chosen locations clearly depend on the POD mode, which was aimed by the optimization; therefore it becomes quite complicated to choose, manually from the extrema of several POD modes, one set of sensors, to reconstruct these POD modes altogether. In this situation the use of the optimization algorithm is particularly helpful and interesting.

The algorithm is used, this time, to found a set of 20 sensors optimized to reconstruct the sum of the 20 first POD modes simultaneously. Fig. 16 compares the determination coefficients of the first 30 POD modes using the grids of 33 and 21 sensors, as well as the set of 20 sensors obtained by the optimization algorithm. The LSE using

the optimized locations leads to a better reconstruction and prediction of the 20 first POD modes compare to the grids of 21 sensors, and can even compete with the grids of 33 sensors. For POD modes of higher rank, grids of 33 sensors seem to perform better, and the grids of 21 sensors and the set of 20 sensors perform comparably. With the optimized set of sensors, the cut-off rank, beyond which POD modes are predicted with less than 10% of R^2 , is about 30. It is an improvement compared to the grids of 21 sensors (14 and 21 modes for “top” and “bottom” respectively). Moreover, the first two POD modes are predicted with R^2 higher than 80%. Therefore, the location of the sensors clearly plays an important role on the length scales that are predicted by the LSE.

At last, the determination coefficient R^2 of the prediction of the entire flow field by LSE with the set of 20 sensors (from the algorithm) accuracy is about 35%, which is higher than the values obtained with 21 and 33 sensor grids, and of the same order of magnitude as with the grid of 56 sensors (see Table 2). This shows the effectiveness of the sensors optimization algorithm, which is able to select a limited number of locations that lead to a better LSE reconstruction and prediction than uniform sensor grids with more sensors.

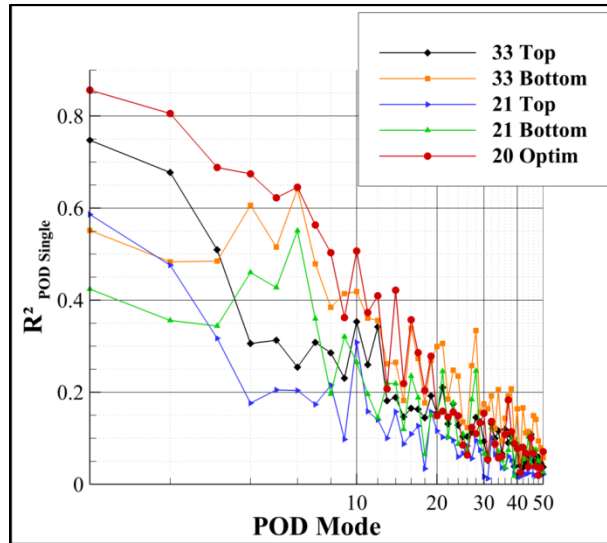


Fig. 16: Determination coefficient as function of the POD mode

VI. Stochastic Estimation flow chart

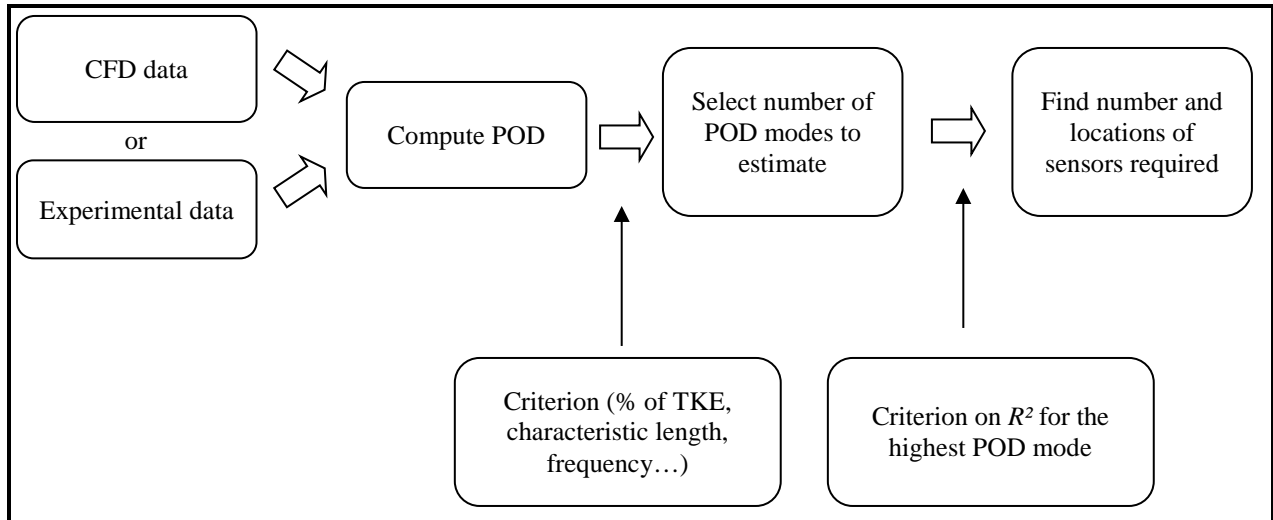


Fig. 17: Flow chart for the use of SE

From the previous results, it is possible to draw a flow chart to estimate turbulent flows thanks to SE (see Fig. 23). The first step would be to compute the POD of the flow, from experimental data or simulated data. Once the

POD obtained and given a criterion (TKE level, characteristic length scales or frequencies), the number of POD modes to estimate is obtained. Then a level of accuracy has to be chosen in terms of R^2 . At last, using this target value of R^2 , the sensor location optimization algorithm (or its extension) is used in order to estimate the minimum number of sensors (or conditional events if synchronized and delayed data are considered), and their locations (and potentially delays), required to match the chosen conditions.

VII. Conclusions

In this paper, the difficulty to estimate the small turbulent spatial integral length scales by Linear Stochastic Estimation (LSE) using only wall pressure measurements has been demonstrated. Fields of turbulent spatial integral length scales show that LSE overestimates these scales even in the areas where they are the largest. The comparative study of several in flow streamwise velocity sensor grids shows that it is possible to estimate smaller length scales by refining the sensor grids. However, if it is possible to reconstruct and predict turbulent integral length scales of half the spacing of coarse grids, for fine grids the smallest length scales recovered are of the order of magnitude of the grid spacing. In a sense, the LSE filters the spatial integral length scales shorter than half the sensors spacing.

The investigation of the Proper-Orthogonal-Decomposition (POD) modes estimation shows that only the first POD modes are correctly estimated and they correspond to the largest scales of the flow in the present case. Refining the grids allows a better estimation of higher order POD modes, but even with very fine grids, only a few POD modes are reconstructed and predicted with a high fidelity. This investigation also highlights the strong importance of the sensors location. It shows that a POD mode is better estimated if the sensors are located close to its extrema. Therefore the possibility to predict with high accuracy a velocity POD mode, from sensors located at the wall only, is limited.

The use of the sensor location optimization algorithm proposed by Muradore et al. [16] has been investigated. The locations obtained by the algorithm are close to the extrema of the POD mode, which is consistent with Cohen et al. [15] conclusion. In addition, the algorithm represents a convenient tool to determine sensors location in order to predict several POD modes simultaneously. It also shows the possibility to estimate smaller length scales, and higher POD modes, with the same amount of sensors with a different placement. Therefore the sensors location plays an important role on the estimation of POD modes and thus on the integral length scales estimated by the LSE.

VIII. References

- [1] R. J. Adrian, "On the role of conditional averages in turbulence theory," *Turbulence in liquids*, pp. 323-332, 1977.
- [2] J. A. Taylor and M. N. Glauser, "Towards Practical Flow Sensing and Control via POD and LSE Based Low-Dimensional Tools," *Journal of Fluids Engineering*, vol. 126, no. 3, pp. 337-345, 2004.
- [3] V. Durguesh and J. W. Naughton, "Multi-time-delay LSE-POD complementary approach applied to unsteady high-Reynolds-number near wake flow," *Experiments in Fluids*, vol. 49, pp. 571-583, 2010.
- [4] J. H. Tu, J. Griffin, A. Hart, C. W. Rowley, L. N. Cattafesta and L. S. Ukeiley, "Integration of non-time-resolved PIV and time-resolved velocity point sensors for dynamic estimation of velocity fields," *Experiments in Fluids*, vol. 54, pp. 1429-1449, 2013.
- [5] H. Clark, A. Naghib and P. Lavoie, "General perspectives on model construction and evaluation for stochastic estimation, with application to a blunt trailing edge wake," *Experiments in Fluids*, vol. 55, no. 7, 2014.
- [6] Z. Hosseini, R. J. Martinuzzi and B. R. Noack, "Sensor-based estimation of the velocity in the wake of a low-aspect-ratio pyramid," *Experiments in Fluids*, vol. 56, no. 1, January 2015.
- [7] D. Lasagna, L. Fronges, M. Orazi and G. Iuso, "Nonlinear Multi-Time-Delay Stochastic Estimation: Application to Cavity Flow and Turbulent Channel Flow," *AIAA Journal*, vol. 53, no. 10, pp. 2920-2935, October 2015.
- [8] L. M. Hudy, A. Naguib and W. M. Humphreys, "Stochastic estimation of a separated-flow field using wall-pressure-array measurements," *Physics of Fluids*, vol. 19, 2007.
- [9] T. D. Nguyen, J. C. Wells, P. Mokhasi and D. Rempfer, "Proper orthogonal decomposition-based estimations of the flow field from particle image velocimetry wall-gradient measurements in the backward-facing step flow," *Measurement Science and Technology*, vol. 21, no. 11, 2010.
- [10] A. Arnault, J. Dandois, J. Delva, J.-C. Monnier and J.-M. Foucaut, "Assessment of Stochastic Estimation

- ability to reconstruct velocity fields," in *50th 3AF International Conference on Applied Aerodynamics*, Toulouse, France, 2015.
- [11] J. L. Lumley, "The structure of inhomogeneous turbulent flows," *Atmosphere turbulence and radio propagation*, pp. 166-178, 1967.
- [12] L. Sirovich, "Turbulence and the dynamics of coherent structures, part1: Coherent structures," *Quarterly of Applied Mathematics*, vol. 45, no. 3, pp. 567-571, 1987.
- [13] N. E. Murray and L. S. Ukeiley, "Estimating the shear layer velocity field above an open cavity from surface pressure measurements," AIAA 2002-2866, in *32nd AIAA Fluid Dynamics Conference and Exhibit*, Saint Louis, 2002.
- [14] P. L. O'Neill, D. Nicolaides and J. Soria, "Autocorrelation Functions and the Determination of Integral Length with Reference to Experimental and Numerical Data," in *15th Australasian Fluid Mechanics Conference*, The University of Sydney, Sydney, Australia, 2004.
- [15] K. Cohen, S. Siegel, M. Luchtenburg, T. McLaughlin and A. Seifert, "Sensor Placement for Closed-Loop Flow Control of a "D" Shaped Cylinder Wake," AIAA 2004-2523, in *2nd AIAA Flow Control Conference*, Portland, Oregon, 2004.
- [16] R. Muradore, F. Bezzo and M. Barolo, "Optimal sensor location for distributed-sensor systems using multivariate regression," *Computers and Chemical Engineering*, no. 30, pp. 521-534, 2005.

