Reconstruction of turbulent velocity fields from punctual measurements

Anthony Arnault

N° d'ordre : 305

Centrale Lille

THÈSE

Présentée en vue d'obtenir le grade de

DOCTEUR

en

Mécanique

par

Anthony Arnault

DOCTORAT DÉLIVRÉ PAR CENTRALE LILLE

Titre de la thèse :

Reconstruction de champs aérodynamiques à partir de mesures ponctuelles

Soutenue le 13 Décembre 2016 devant le jury d'examen :

Président du jury	Laurent-Emmanuel Brizzi	Professeur, Pprime, Poitiers
Rapporteur	Bernd Rainer Noack	Directeur de recherche au CNRS, LIMSI, Orsay
Membre	Jean-Camille Chassaing	Maître de Conférence, IJLRA, Paris
Membre	Bérengère Podvin	Chargé de recherche au CNRS, LIMSI, Orsay
Invité	Lionel Mathelin	Chargé de recherche au CNRS, LIMSI, Orsay
Encadrant ONERA	Julien Dandois	Ingénieur de recherche à l'ONERA, Meudon
Directeur de thèse	Jean-Marc Foucaut	Professeur à l'Ecole Centrale Lille, Lille

Thèse préparée dans le Laboratoire de Mécanique de Lille et au Département d'Aérodynamique Appliquée de l'ONERA

Ecole doctorale SPI 072 (Centrale Lille)

Remerciements

I would first like to thank Prof. Bernd R. Noack for kindly reviewing my work. His understanding and challenging vision on estimation and flow control are inspiring.

Je remercie aussi vivement le Professeur Laurent-Emmanuel Brizzi d'avoir bien voulu rapporter mes travaux et de m'avoir fait l'honneur de présider mon jury de thèse.

Je tiens à remercier Madame Bérengère Podvin et Messieurs Jean-Camille Chassaing et Lionel Mathelin pour l'examen de ces travaux et l'intérêt qu'ils y ont porté.

La direction de ses travaux a été assurée par le Professeur Jean-Marc Foucaut dont les riches échanges ont permis de correctement jalonner ces trois années.

Je remercie très chaleureusement Monsieur Julien Dandois, tout d'abord de m'avoir accueilli en stage et ensuite d'avoir bien voulu de moi en thèse malgré ma « mémoire de poisson rouge ». Son implication, sa disponibilité, ainsi que son écoute et son humour caustique, ont été essentiels à la tenue de ses travaux et m'ont été très utiles dans les moments difficiles.

Je souhaite remercier les membres du projet HYBEXCIT dans lequel s'inscrit ma thèse, Madame Anne Gilliot et Messieurs Jérôme Delva, Jean-Claude Monnier, Quentin Gallas, Bruno Mialon, en particulier pour la réalisation des cas tests expérimentaux utilisés dans ses travaux, et aussi pour leur sympathie lors de mes passages au centre ONERA de Lille.

Cette thèse a été réalisée au sein du département d'aérodynamique appliquée (DAAP) sur les centres ONERA de Lille et de Meudon. Je remercie tous mes collègues Lillois, comme Meudonnais, pour le cadre de travail très agréable et les échanges intéressants qui ont participé à mon enrichissement professionnel et personnel. Merci à Brigitte Duda (pour sa bonne humeur et ses beaux albums photos), à Murielle Lamoureux, à Christophe Verbecke, à Jean-Michel Desse et à Jean-Paul Bourrez (et son punch !). Je remercie également François Olchewsky d'avoir partagé son bureau et de m'avoir détendu lors de mon combat avec Orange ! Merci aussi à Antoine Dumont, Fabien Gand, Michel Costes, Didier Bailly, Daniel Desterac. Un grand merci aux joueurs du midi Jérémy Duchemin, Jean-Luc Hantrais-Gervois et en particulier à Christelle Wervaecke pour son soutien et sa precieuse écoute. Je remercie Gérald Carrier et Joël Reneaux de m'avoir accueilli dans leur service. Une pensée pour les doctorants avec qui j'ai eu l'occasion de passer du temps et de partager ces années : Aurélien Arntz, Hélène Toubin, Jonas Verrière, Andrea Viti, Illias Petropoulos, Amaury Bannier, Nicolas Renard et bien d'autres.

Un grand merci à mes amis qui m'ont accompagné durant ces années, voire jusqu'à la soutenance pour certains : Thomas Henriat (notamment pour les billets Disney…mais pas que !), Marine Richard, Florian Wintz, Mathias Fages, Alexandre Pasquier, Elise Massot, Laetitia Leduc, Aude Leick (et j'en oublie mais j'ai une mémoire de poisson rouge !). En particulier, je souhaite remercier Lucien Xu pour les parties de jeux jusque tard le soir parfois, et surtout pour ces précieuses connaissances informatiques.

Je remercie les membres de ma famille pour leurs encouragements tout au long de ces années et pour leur confiance.

Enfin, merci à Nicolas Desplan de partager chaque jour de ma vie, même les plus mauvais, et d'avoir bien voulu sponsoriser cette aventure !

Résumé

Cette thèse vise à étudier et à développer des outils afin d'estimer, en temps réel, des champs de vitesse d'écoulements turbulents à partir d'un nombre limité de mesures ponctuelles. L'estimation stochastique est en particulier examinée dans ces travaux.

Le premier chapitre consiste en une étude bibliographique de l'estimation stochastique et des méthodes qui lui sont dérivées. Le développement chronologique de l'estimation stochastique est d'abord présenté. En particulier, les différentes extensions de l'estimation stochastique sont ainsi décrites dans l'ordre de leur apparition. Dans un second temps, les applications de l'estimation stochastique à la prédiction en temps réel d'écoulement sont passées en revue. Cette étude bibliographique a permis de mettre en avant l'étendue de l'utilisation de l'estimation stochastique afin d'étudier la physique d'écoulements turbulents. L'utilisation de l'estimation stochastique en vue de fournir des prédictions de champs de vitesse est par contre plus récente. Les résultats obtenus sont encourageants mais concernent principalement des écoulements relativement simples dont la dynamique peut être décrite avec un faible nombre de modes POD (Décomposition en modes propres orthogonaux). Cette étude bibliographique a aussi montré l'absence d'une méthodologie commune afin d'évaluer la précision des prédictions obtenues par estimation stochastique.

Le deuxième chapitre fournit une description mathématique de l'estimation stochastique et de ses extensions. La décomposition en modes propres orthogonaux est aussi décrite. Le choix des écoulements tests, qui serviront à évaluer les capacités de l'estimation stochastique à fournir une prédiction précise des champs de vitesse d'écoulements turbulents, est aussi discuté. Les différentes bases de données, numériques et expérimentales, correspondantes aux différents écoulements sélectionnés sont ensuite décrites. De plus, la convergence des statistiques utilisées par l'estimation stochastique est en particulier analysée.

Les capacités de l'estimation stochastique (et de ses méthodes dérivées) à fournir des prédictions des champs de vitesse d'écoulements turbulents sont évaluées dans le chapitre 3. L'estimation stochastique et ses extensions sont utilisées pour prédire les champs de vitesse de plusieurs écoulements de complexité croissante. Dans tous les cas testés, les prédictions les plus précises sont obtenues en utilisant l'estimation stochastique à plusieurs délais temporels (avec régularisation de Tikhonov). L'utilisation de l'estimation stochastique couplée à la POD permet avant tout de réduire la taille du problème à résoudre, mais n'améliore pas particulièrement la précision de la prédiction. Entre l'estimation stochastique linéaire et quadratique, l'estimation quadratique semble pouvoir mener à des prédictions plus précises, mais est aussi plus sujette au problème de sur-apprentissage. Enfin, cette évaluation a mis en avant de fortes limitations de l'estimation stochastique. Si les prédictions obtenues dans certains cas tests sont très précises, elles le sont beaucoup moins dans d'autres cas tests.

Bien que ces observations ne soient pas une entière surprise, le chapitre suivant est consacré à caractériser plus précisément l'estimation stochastique afin de mieux expliquer les raisons qui mènent à ce qu'un écoulement soit mieux reconstruit qu'un autre. Tout d'abord l'impact d'un bruit Gaussien est évalué. Cette évaluation a montré que les performances de l'estimation stochastique sont sous-estimées lorsqu'elles sont déterminées en utilisant des données expérimentales. En particulier, le bruit présent dans les mesures servant à la reconstruction a un très fort impact sur la qualité de la prédiction. Les estimations de champs de vitesse numériques, d'un même écoulement, mais issus de deux méthodes de simulations différentes sont ensuite comparées. L'une des simulations est une Unsteady Reynolds Averaged Navier-Stokes (URANS) simulation. Dans cette simulation, toute la turbulence de l'écoulement est modélisée. La deuxième simulation est une Zonal Detached Eddy Simulation (ZDES) dans laquelle une part de la turbulence de l'écoulement est directement

simulée. La comparaison des estimations de ces deux simulations permet d'observer l'impact de la turbulence sur la qualité de l'estimation. De cette comparaison, il ressort que le contenu haute fréquence et de courte échelle intégrale de la turbulence est mal prédit par l'estimation stochastique. Suivant ces observations, la prédiction des échelles intégrales de la turbulence est analysée. Cette étude a montré que les grandes échelles de l'écoulement sont les mieux prédites par l'estimation stochastique. Cette analyse a aussi mis en avant le caractère filtrant de l'estimation stochastique. Suivant la position des capteurs utilisés différentes échelles peuvent être reconstruites. En particulier, l'étude de l'estimation des modes POD a montré que la prédiction d'un mode est plus précise lorsque des capteurs sont positionnés à proximité des nœuds du mode POD. De plus, il a aussi été montré qu'il est possible de définir une longueur caractéristique d'un mode POD, correspondant à la plus petite structure contenue par ce mode POD et qu'un mode POD correspondant à une courte échelle caractéristique possède plus de nœuds qu'un mode de grande échelle caractéristique. Ainsi pour prédire correctement un mode de courte échelle caractéristique, il faut davantage de capteurs que pour prédire un mode POD ayant une grande échelle caractéristique. Ces observations mettent très clairement en avant le fort impact de la position des capteurs sur la prédiction. Le caractère filtrant de l'estimation stochastique concernant les échelles contenues dans l'écoulement a ensuite été observé en étudiant l'impact d'un pré-filtrage spatiale des champs de vitesse à estimer. L'étude du pré-filtrage temporel des données à estimer et servant à l'estimation a aussi permis de confirmer que le contenu haute fréquence d'un écoulement turbulent était prédit avec une plus faible précision que le contenu basse fréquence. Enfin l'impact du nombre de Reynolds sur la qualité de la prédiction par estimation stochastique d'un même écoulement a été évalué. L'un des effets principaux de la diminution du nombre de Reynolds est la concentration de l'énergie cinétique turbulente de l'écoulement dans un plus faible nombre de modes POD. En d'autres termes, un écoulement faiblement turbulent contient davantage d'énergie dans les grandes structures de l'écoulement qui sont correctement reconstruites, ce qui explique qu'il soit prédit avec une meilleure précision qu'un écoulement plus turbulent qui contient davantage d'énergie dans les petites structures.

De ces caractérisations, il est apparu qu'un moyen évident d'améliorer la qualité d'une prédiction par estimation stochastique est d'optimiser les évènements conditionnels utilisés (i.e. la position des capteurs et leurs délais temporels si l'estimation stochastique à plusieurs délais est employée). A cette fin, un algorithme d'optimisation de la position des capteurs a été testé et sa capacité à choisir des positions de capteurs améliorant la précision des prédictions a été confirmée (chapitre 5). Cet algorithme a ensuite été étendu aux choix de la position des capteurs et de délais temporels. Il a été montré que cet extension de l'algorithme permet de choisir un faible nombre de délais menant à la même qualité de prédiction que lorsqu'un plus grand nombre de délais répartis régulièrement dans le temps. Néanmoins la précision de la prédiction n'a pas pu être améliorée en utilisant cet algorithme étendu.

Les résultats obtenus dans les précédents chapitres ont montré que l'estimation stochastique ne permet pas d'obtenir des prédictions hautement précises pour certains cas tests. Il a donc été décidé de tester dans le chapitre 6, une autre méthode permettant de produire des prédictions de champs de vitesse d'écoulements turbulents à partir de mesures ponctuelles, le filtre de Kalman. Le filtre de Kalman permet de prendre en compte la dynamique de l'écoulement en utilisant un modèle dynamique. Dans ce chapitre, le filtre de Kalman a été testé sur deux cas tests (l'écoulement autour du profile OAT15A et autour d'un cube) en utilisant un ensemble de modes POD comme état. Les résultats obtenus ont montré qu'il était possible d'obtenir une amélioration substantielle de la précision de la prédiction en utilisant le filtre de Kalman étendu, avec un modèle d'observation quadratique, dans le cas de l'OAT15A, comparé aux meilleures prédictions obtenues avec l'estimation stochastique linéaire à plusieurs délais. Les résultats obtenus avec l'écoulement autour du cube sont moins encourageant, et aucun des filtres de Kalman testés n'a permis d'améliorer la qualité de la prédiction en comparaison de l'estimation stochastique linéaire à plusieurs délais.

Nomenclature

t	Time
f	Frequency
$\boldsymbol{u}$	Fluctuating velocity vector
u	Streamwise component of the velocity
v	Vertical component of the velocity
w	Spanwise component of the velocity
P	Static pressure
ν	Kinematic viscosity
ρ	Density
E	Conditional event
a	POD Coefficient
Φ	POD mode or RBF
Ω	Spatial domain
h	Height
c	Chord
Re	Reynolds number
St	Strouhal number
R^2	Determination coefficient
δ	Boundary layer thickness
δ^*	Boundary layer displacement thickness
θ	Boundary layer momentum thickness
λ	Characteristic length scale for POD mode
Λ	Turbulence spatial integral length scale

Acronyms

SE	Stochastic Estimation
LSE	Linear Stochastic Estimation
QSE	Quadratic Stochastic Estimation
HOSE	Higher Order Stochastic Estimation
MTD	Multi-Time-Delay
POD	Proper Orthogonal Decomposition
EPOD	Extended Proper Orthogonal Decomposition
FFT	Fast Fourier Transform
CFD	Computational Fluid Dynamics
RANS	Reynolds Averaged Navier-Stokes
URANS	Unsteady Reynolds Averaged Navier-Stokes
(Z)DES	(Zonal) Detached Eddy Simulation
BFS	Backward Facing Step
TKE	Turbulent Kinetic Energy
PSD	Power Spectral Density
SNR	Signal-to-Noise Ratio
KF	Kalman Filter
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
RBF	Radial Basis Function
(LOO)CV	(Leave One Out) Cross Validation
ODE	Ordinary Differential Equation

Contents

Introduction

Obtaining accurate flow field information is a critical and more and more important necessity in several areas. For instance, the monitoring of pollutant dispersion in urban area has become a major issue in the past decade. The growth of air traffic has also led the European Transport Safety Council to push for better security measures in European airport [1]. One of them concerns wake vortex constraints which govern the minimum required distance between aircraft lined up in sequence on the approach to the runway. The council stressed the need for research on modeling wake vortex and on methods to identify and locate wake vortices from ground based or airborne sensors.

The development of shipboard helicopter operations is also strengthening the need for precise aerodynamic information around the flight deck. Indeed, shipboard operations are among the most difficult manoeuver for helicopters. The main causes that make such operations so challenging are the ship motion and its air wake [2]. In order to make these operations safer operational flight envelopes are defined for every couple of helicopter/ship, forming the Ship-Helicopter Operating Limits (SHOL). These flight envelopes have been an effective way to secure shipboard operations. However, they limit the operational conditions. An extreme example is presented by Healey [3] who stated that "a helicopter can operate a 125 m (400 ft.) frigate in the North Sea a mere ten percent of the time in winter". Thus efforts are made to improve the security of shipboard operations without extending the SHOL. One approach is to decrease turbulence levels and side and down winds over the helipad. As a first step, it is necessary to have precise flow information around the flight deck. That information can then be used to apply flow control or be directly sent to the helicopter in order to be integrated in its control. In a general way, the control of turbulence first requires a correct estimation of the state of the flow.

In this context, ONERA has launched a Research Project called HYBEXCIT in 2013. The project aims to develop methods for the reconstruction of 3D aerodynamic field from numerical and experimental data using sparse measurements.

A class of methods that have received much attention in turbulent flow and recently been used to estimate aerodynamic field is the Stochastic Estimation (SE). The SE was first introduced by Adrian [4] [5] and Adrian et al. [6] for turbulent flows to educe coherent structures from a few sparse conditional events. They showed that the SE can approximate conditional averages of turbulent flow quantities. They demonstrated that the SE is a simple tool to detect coherent structures and determine, qualitatively, their properties. Following their work, the SE has been applied to a large range of turbulent flows, such as: axisymmetric jet shear layer [7], cylinder wake [8], plane shear layer [9], flat-plate turbulent boundary layer [10], cavity flow [11], detached flow [12] [13]. These numerous applications of the SE proved the usefulness to get a better understanding of turbulent flow dynamics.

Since its introduction, the SE has also been developed in different ways. The Linear SE and Quadratic SE (QSE) (as well as more generally Higher-Order SE) have been investigated and compared early on by Tung et al. [14] and Guezennec et al. [15]. Both studies concluded that differences between LSE and QSE were marginal. Later on, Naguib et al. investigated the use of SE to estimate velocity from pressure measurements and concluded that the inclusion of the quadratic term was necessary to obtain a representative estimation of the conditional average [16]. The SE has also been coupled with the Proper-Orthogonal-Decomposition (POD). Bonnet et al. first proposed to project the flow field, obtained by SE, onto the POD basis [17]. This method is called the complementary technique. They showed that the phase information of the POD modes could be retained by the complementary technique. Another coupling has been proposed by Taylor et al. [12]. In their coupling, the POD coefficients, of the decomposition of the flow, are directly estimated using SE.

This coupling is called the modified complementary technique, or simply SE-POD. The advantage of this coupling is that by selecting a few POD modes to estimate, the problem dimension is greatly reduced. The SE-POD is the more commonly used POD and SE coupling in recent works. At last, Multi-Time-Delay SE, in spectral and temporal domain, has been developed. Tinney et al. [18] described the spectral SE and showed its usefulness when the spectral features of the measured and estimated data are disparate and/or significant time delay exists between them. The spectral SE has been applied mostly for jet flows and is limited to offline applications as stated by Sinha et al. [19]. In the temporal domain, the MTD-SE has been precisely described by Durgesh et al. [20] who showed that the accuracy of the estimate was improved using MTD-SE instead of single-time SE.

While the SE has been extensively used in order to study the physics of turbulent flows, its application to real time flow prediction, from a limited number of sensors, is more recent. Such application was clearly identified for the first time by Taylor et al. [12] in 2004. They investigated the ability of the LSE-POD to accurately estimate the velocity field from wall pressure measurements. Their final objective was to develop an active feedback control for separated flows. Following their work, several authors used the SE, and often the MTD-LSE-POD, to build Reduced-Order-Model (ROM) to be used in flow control. In particular, Durgesh et al. and Tu et al. showed the ability of MTD-LSE-POD to estimate with high accuracy the velocity field. However, such applications of SE are limited to relatively simple turbulent flows, for which the dynamic can be described with a few POD modes (less than 10). Results of Taylor et al. concerned a backward facing ramp at a Reynolds number based on the ramp height (Re_h) of 2×10^4. Rowley et al. [21] used a cavity flow at $Re_d = 395$ (based on cavity depth d), Nguyen et al. a backward facing step at $Re_h = 2.8\times10^3$ [22], Durgesh et al. the wake downstream a bluff body at $Re_h = 2\times10^4$ and Tu et al. a similar flow at $Re_h = 3.6\times10^3$.At last only a few recent studies truly investigated and evaluated the accuracy of SE to predict turbulent flows [22] [23].

Lastly the use of Kalman Filter (KF), coupled with ROM based on POD Galerkin projection, has also been proposed. Mokhasi et al. [24] showed promising results of real-time prediction of the flow around a cube mounted in a channel ($Re_h = 4\times10^4$) using unscented KF. Sinha et al. compared LSE, QSE and KF in an axisymmetric jet [19]. They observed that the KF and QSE outperform the LSE.

To sum up this short literature overview, the SE has been used for several decades in turbulence as an efficient and simple tool to study coherent structures. But its application to real-time flow prediction is much more recent. First works concerning this application have shown the attractiveness of the technique (in particular of the MTD version) and demonstrated its performance on simple or academic turbulent flows. Therefore, the application of SE in more challenging and realistic situations remains undone. As an alternative to the SE, the Kalman Filter has also been applied. While the Kalman Filter is a more computationally expensive method, it has been shown that its performance, for real-time flow prediction, was better than the one of SE.

Hence this study which consists in investigating SE methods and developing tools in order to produce real-time estimation of the velocity field of highly turbulent flows from a few punctual measurements.

The first chapter reviews the state of the art concerning SE methods and highlights some points that should be more investigated.

The second chapter described most of the mathematical background of the methods used in this work. In addition, the experimental and numerical databases that are used in order to test the estimation methods are detailed. In particular, the choice of the test cases is justified and special care is given to the convergence of the statistical moments used in SE techniques.

Then, SE methods are applied to the several flow configurations previously selected in order to evaluate their ability to predict, in real-time, highly turbulent flows from a limited number of measurements. (Chapter III)

Since the quality of the SE reconstruction is quite limited for some of the flow configurations tested, it was decided to better characterize the method. First, the impact of Gaussian noise in the data to estimate and in the measurements is studied. Then, estimations through SE of a same flow using two different numerical data sets, one from an Unsteady Reynolds Averaged Navier-Stokes (URANS) simulation and one from a Zonal Detached Eddy Simulation (ZDES). In URANS, the turbulence is entirely modeled whereas in ZDES part of the turbulence is directly resolved and not modeled. Therefore, the comparison of estimation from both simulations can be used to identify the impact of the turbulence on the SE methods accuracy. Following this characterization, the estimation of the turbulent spatial integral length scales of the velocity field is then investigated. The filtering effect of SE is highlighted and the impact of the conditional events used is also particularly emphasized. The effect of prefiltering the measurements and/or the training data spatially or temporarily is then studied in an attempt to better characterize the filtering effect of the SE. At last, the impact of an increase of Reynolds number on the SE reconstruction accuracy is investigated. (Chapter IV).

Following the study of the turbulent spatial integral length scale estimation, which demonstrated the strong influence of the sensor locations on the estimation accuracy, the chapter V devoted to the development and test of an optimization algorithm for choosing the locations of the conditional events. The algorithm is used to determine a strategy of utilization of the SE-POD. Then this optimization algorithm is extended to the choice of delays for the MTD-SE. (Chapter V)

Finally, the Kalman Filter and several of its extensions are used to produce real-time flow prediction in some of the flow configurations employed to test the SE. Several strategies to obtain the dynamic model and measurement model required by KF are compared. In all cases, a ROM formed with POD is used and the predictions obtained are compared with those of the SE methods. (Chapter VI)

Communications

The present work has been presented in two conferences and part of it has also been published.

Conference Proceedings

A. Arnault, J. Dandois, J.-C. Monnier, J. Delva and J.-M. Foucaut, “Assessment of Stochastic Estimation ability to reconstruct velocity fields”, in *50th International Conference on Applied Aerodynamics*, Toulouse, France, 2015.

A. Arnault, J. Dandois, J.-C. Monnier, J. Delva and J.-M. Foucaut, “Analysis of the filtering effect of the stochastic estimation and accuracy improvement by sensor location optimization”, in *8th AIAA Flow Control Conference*, 2016, Washington D.C., AIAA 2016-3257.

Peer-reviewed journal articles

A. Arnault, J. Dandois and J.-M. Foucaut, “Comparison of stochastic estimation methods with conditional events optimization for the reconstruction of the flow around a supercritical airfoil in transonic conditions”, *Computers and Fluids*, vol. 136, pp. 436-455, 2016.

A. Arnault, J. Dandois, J.-C. Monnier, J. Delva and J.-M. Foucaut, “Analysis of the filtering effect of the stochastic estimation and accuracy improvement by sensor location optimization”, *Experiments in Fluids*, vol. 57, no. 12, pp. 185-207, 2016.

Chapter I. Literature review on Stochastic Estimation

The aim of this part is to present a literature overview of the Stochastic Estimation (SE) which has led to mainly investigate the use of SE to get accurate estimate of turbulent flows. A separate short literature review focusing on Proper-Orthogonal-Decomposition Reduced-Order-Model (POD-ROM) and data assimilation is given at the beginning of Chapter VI.

The following literature review is organized so that the chronological development of SE in the field of turbulence are described in a first section. The several extensions of SE are in particular described. Then, in a second section, the application of SE in order to produce real-time flow prediction are reviewed.

1. Developments of the Stochastic Estimation

The Stochastic Estimation (SE) was first introduced in the field of turbulence by Adrian in the late 70's [4] [5] and Adrian et al. [25] [6]. They showed that conditional averages of turbulent flow quantities can be approximated in terms of unconditional correlation data using SE. In the SE, the conditional average is approximated through a Taylor expansion around the average value of the conditional event E (for example punctual measurements of pressure, velocity, skin friction…). The SE is expressed by the following formula (where u is a zero-mean quantity, N is the number of y_i points where E is known):

$$\tilde{u}(\boldsymbol{x},t) = \sum_{i=1}^{N} \mathrm{A}_{\boldsymbol{x}}(y_i)\,\mathrm{E}(y_i,t) + \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{B}_{\boldsymbol{x}}\left(y_i,y_j\right)\mathrm{E}(y_i,t)\mathrm{E}(y_j,t) + \cdots$$

The coefficients $\mathrm{A}_{\boldsymbol{x}}(y_i)$, $\mathrm{B}_{\boldsymbol{x}}\left(y_i,y_j\right)$ are computed using a training data set in which the true $u(\boldsymbol{x},t)$ and the conditional events $\mathrm{E}(y_i,t)$ are known. More details on the method are given in Chapter II.1.

Adrian et al. applied the SE, using experimental data, in four turbulent flows: grid turbulence, the axisymmetric shear layer of a round jet, a plane shear layer and a pipe flow. In their application, the approximation, by SE, of conditional average of the velocity at a location given the unconditional velocity at another location was investigated. They concluded that SE can be used to extract, in turbulent flows, coherent structures characterized quantitatively by a conditional eddy. For instance, Adrian deduced from the Linear Stochastic Estimation (LSE) of a shear flow the presence of hairpin vortex. Thus the SE was introduced, in the field of turbulence, as a tool to detect coherent structures and determine quantitatively their properties. The main goal was then to get a better understanding of the flow physics. A first review of several applications of the LSE to educe coherent structures in turbulent flows was given by Adrian in 1994 [26]. A deep presentation of the SE, describing its fundamental concepts and applications to the extraction of conditional eddy structures, was also given by Adrian in [27].

1.1. Single-point SE vs. multipoints SE

As a tool to study coherent structures in turbulent flows, SE has been developed in several directions. In the first applications of SE by Adrian, only one measurement (at one location) was

utilized. Very quickly, the use of additional events was investigated. In 1989, Guezennec [15] used two conditional events (two velocity measurements) to study the structures in a flat plane boundary layer. Experimental data were used and the Reynolds number based on the momentum thickness in the plane of measurements was 4900. Using the SE, he was able to observe convection and deformation of structures in the flow. In 1992, Cole et al. [7] applied the SE to an experiment of a jet mixing layer. Quantifying the results in term of TKE captured by the SE, they found that the use of a single point velocity measurements leads to an inadequate representation of the instantaneous velocity field. On the contrary, the use of two points velocity measurements located on opposite sides of the shear layer yielded more realistic estimates. The addition of more velocity measurements was also studied and showed little gain compared with the two points case. Giezeke et al. [8] then applied multi-point LSE to examine the three-dimensional structure of a turbulent wake of a cylinder using experimental data. In addition, they performed a pseudo-dynamic reconstruction of the flow field. Later, the impact of the location and number of unconditional events was also studied by Vincendeau [9] using an experiment of a shear layer. In his work, Vincendeau investigated three-dimensional structures in the flow and in particular the large scale structures matching using LSE. He showed that different level of TKE could be recovered with the same number of velocity probes but at different locations. He concluded that more TKE was restored when velocity measurements were closer to the center of the mixing layer. However large coherent structures were better identified when the velocity measurements were located further on each side of the mixing layer. Recently, Dekou [28] investigated the large scale structures developing in a wall turbulent flow. LSE was used to reconstruct the three component of the velocity field using a grid of 143 hot wire sensors. They showed that the LSE satisfactorily retained the information on the large structures which study was aimed.

1.2. Quadratic Stochastic Estimation and Higher-Order Stochastic Estimation

From the beginning Adrian et al. [14] also studied the impact of Higher Order Stochastic Estimation (HOSE), especially Quadratic Stochastic Estimation (QSE) in the case of isotropic turbulence. The use of QSE is also investigated by Guezennec [15] in turbulent boundary layer. In QSE or HOSE, terms of order 2 or higher are conserved in the Taylor expansion used to approximate the conditional averages. At the time both studies concluded that improvements of accuracy between HOSE and LSE, in the estimation of the conditional averages, remain marginal and did not warrant the additional cost of computing higher-order statistical moments. However, in 1992, Brereton [29] provided a framework for assessing the accuracy of SE in predicting conditional averages. Studying the SE applied to a turbulent boundary layer, using only one velocity measurement, he pointed out that the addition of the quadratic term could lead to better estimate the conditional average. To improve the SE, Brereton also proposed to use Laurent series expansion, which contains negative power of the measurements, to approximate the conditional average.

Later on, several researchers investigated the use of other kinds of unconditional events than the velocity, such as pressure or shear stress. In 1988, Adrian [25] already used the velocity as well as the deformation tensor at the same location as measurements. Deformation tensor was also used, as unconditional event, by Olsen et al. [30] who studied the large scale structures in an incompressible shear layer through LSE. In particular, they calculated conditional roller and braid structures from typical values of the deformation tensor corresponding to these structures. Picard et al. [31] applied the LSE to investigate the noise emission mechanism in a subsonic round jet using pressure fluctuations surrounding the jet as unconditional events. They focused on vertical structures developing in the jet and that they found to be responsible for the far-field noise emission. A similar approach has been more recently operated by Gutmark et al. [32] for the study of coherent structures of a swirling jet. Oppositely, Druault et al. [33] investigated the noise production in a 2D compressible plane mixing layer using QSE to estimate the far-field pressure from velocity measurements.

Nicoud et al. [34] studied the use of LSE to generate wall stress model for Large Eddy Simulations (LSE). In his work the wall stress was estimated by LSE using local velocity field as unconditional events.

The use of wall measurements, especially pressure, has also been considered and stimulated by the necessity to simplify experiment campaign and/or the desire to use nonintrusive measurements. In 2001, Naguib et al. [16] studied the ability of SE to capture the conditional streamwise velocity field associated with wall-pressure events in a turbulent boundary layer. They show that the utilization of the wall-pressure, as unconditional events, requires the inclusion of the quadratic term to provide an accurate representation of the conditional velocity field. More precisely, when the Joint Probability Density Functions (JPDF) between the estimated variables and the conditional events are Gaussian, then quadratic and higher-order terms in the Taylor expansion can be neglected. Their results where then confirmed in other studies. For instance, Murray et al. [11] [35] [36] applied LSE and QSE in the case of a cavity flow using wall-pressure measurements. They showed that integral quantities such as the Turbulent Kinetic Energy (TKE) and root-mean-square (RMS) velocity were better estimated by the QSE (see Fig. I.1). Later, they exclusively used QSE to evaluate flow structures in their cavity flow [37] [38]. Such observations, on the performance of QSE, were also reached by Carabello et al. [39] and by Arunajatesan et al. [40] also for cavity flows. Hudy et al. [13] studied a backward facing step flow using LSE and QSE with wall pressure measurements. They showed that the reconstructed structures were similar with both methods (see Fig. I.2). The results were somehow enhanced by the use of QSE. At last, to study differences of flow structures between a baseline axisymmetric jet flow and a chevron jet flow, Kastner et al. [41] used LSE and QSE. They also found that higher levels of TKE were recovered by the QSE.

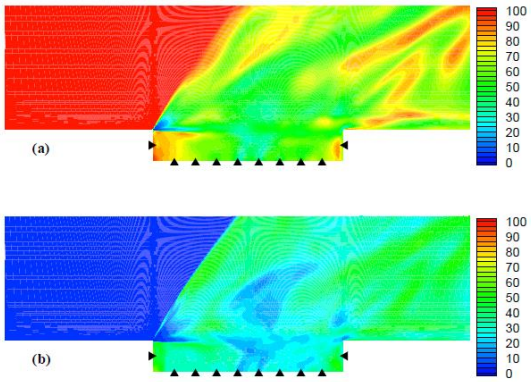


Fig. I.1: Percent of TKE deficit. a) LSE, b) QSE [11]

1.3. Stochastic Estimation coupled with Proper-Orthogonal-Decomposition

Another important development of SE is its coupling with the Proper-Orthogonal-Decomposition (POD). The POD was proposed, in fluid mechanics, by Lumley [42] as an approach based only on an energetic criterion in order to detect coherent structures in turbulent flows. The goal was then to overcome the problems caused by conditional measurements techniques. With the POD, Lumley proposed to consider that the coherent structures are the ones with the largest projection, in the least square sense, on the velocity field. POD has been used to understand the coherent structures in various turbulent flows. [43] in shear layer, by Citriniti et al. [44] in axisymmetric jets, by Johansson et al. [45] in axisymmetric wake and by Cruz et al. [46] in three-dimensional wakes. In one of his first study on SE [5], Adrian shortly discussed the differences between the conditional aver-

ages and the POD, stressing that they are not equivalent but may lead to similarities between extracted eddies. [47] investigated more precisely the link between POD and LSE. In particular, they demonstrated that, under the assumption that the coefficients in the POD are jointly Gaussian, LSE and POD are equivalent. Without that assumption the relationship between the two methods remains difficult to interpret. POD and SE have been used separately in several studies already mentioned above such as those of Picard et al. [31], Murray et al. [37] and Druault et al. [33].

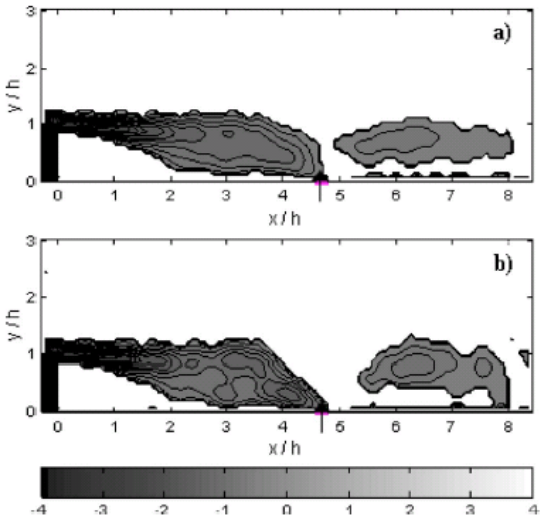


Fig. I.2: Spanwise vorticity field the single-point SE results for a positive pressure events ($p = 5p_{rms}$) (a) LSE and (b) QSE [13]

The idea to couple the LSE and the POD has been first proposed by Bonnet et al. [17] in 1994. In the so-called complementary technique, the LSE of a flow field is projected onto the POD bases in order to obtain estimates of the POD coefficients. The estimated POD coefficients are then used to study the temporal evolution of the first POD mode. They showed that the phase information of the POD modes could therefore be obtained from measurements at a few selected locations. Results obtained for the axisymmetric jet are displayed in Fig. I.3. In 1998, in a joint paper, Bonnet et al. [43] compared several methods to identify structures in turbulent shear flows. Among the methods investigated are the POD, the SE and the complementary technique. Once more they highlighted that it may be possible to obtain time dependent information from the POD and thus the phase of the structures, without measuring the instantaneous velocity at all points in space used to compute the POD. The complementary technique was also investigated by Perret [48] as a part of a method to generate unsteady inlet conditions for LES. Later on, Spitler et al. [49] used the complementary technique in order to estimate, from sparse set of inflow data, the entire wind inflow field of a wind turbine. In 2004, Taylor et al. [12] proposed to directly estimate the POD coefficients through LSE. The technique is called modified complementary LSE or just modified LSE (mLSE). Some papers also refer to this method as LSE-POD (or QSE-POD for the quadratic version). This version of the coupling of POD and SE is the more currently used in recent works [36] [40] [36] [22] [50]. An example of estimation through QSE-POD is shown in Fig. I.4.

Another method exploiting POD and related to SE is the Extended POD (EPOD) which was proposed in 2003 by Borée [51] as a new technique to analyze correlated events in turbulent flows. The idea behind the EPOD is to project the POD coefficients of one POD basis (of a quantity A) onto another set of quantities (B) such as to obtain extended POD modes for B. Borée demonstrated that the EPOD is then the only part of B correlated with A. In addition, he showed that, if all the initial POD modes are used, then the EPOD is equivalent to the LSE, where B is estimated using A as measurements. The EPOD can therefore be used to decompose the LSE. In 2006, Hoarau et al. [52] studied the flow downstream a forward facing ramp. Using the EPOD they extract the part of

the surface pressure signal downstream the reattachment point which is correlated with the pressure signal under the separated region. They showed that the correlated part carries the properties of distorted mixing layer eddies.

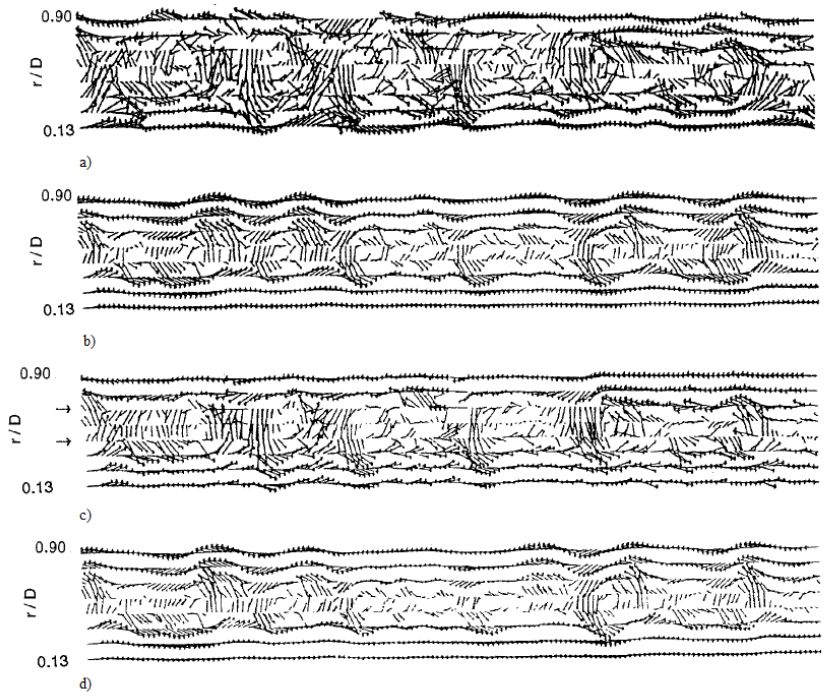


Fig. I.3: Comparison of instantaneous velocity vector plot of a jet seen in a frame of reference moving at U_c = 12 m.s^{-1}. a) Original data, b) projection on the first POD mode, c) LSE estimation using two velocity vector as unconditional events, d) complementary technique using 1 POD mode [17]

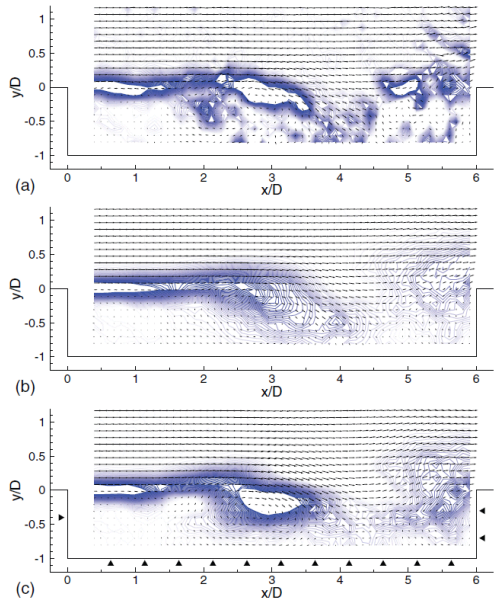


Fig. I.4: Reconstruction of a PIV Snapshot from wall pressure measurements. (a) PIV snapshot. (b) PIV snapshot projected on the first 8 POD modes. (c) Reconstruction by QSE-POD. The contours outline regions of large clockwise rotating spanwise vorticity. The vector scaling and contour levels are the same in each subplot. [36]

1.4. Multi-Time-Delay Stochastic Estimation

Another significant extension of the SE concerns the use of delayed measurements. Early on, a single-time-delay SE (STD-SE) was developed by Guezennec [15]. At the time he uses the velocity measurements at an instant *t* to estimate the velocity field at several others instants. Thus, he obtained a time evolution of the conditional flow using one instantaneous measurement. Through this utilization of the SE, he showed that the convection and deformation of structures could be observed. STD-SE was also used by Naguib et al. [16] who showed its usefulness when a predetermined convective delay could be considered between the conditional measurement and the estimate. In their work of 1999, Ewing et al. [53] applied the complementary techniques of Bonnet using single and multi-time information in the axisymmetric shear layer. In the multi-time configuration they Fourier transform the signal in time, thus using all the time samples as delays. Without calling it so, they therefore introduce the spectral SE. They demonstrated that the method was able, under some conditions on the conditional events, to correctly reproduced the topology of ring structures and inter-ring regions in the jet flow. The spectral SE was later described in detail by Tinney et al. [18] who also applied the technique to an axisymmetric jet flow. They showed the usefulness of the spectral SE whenever the spectral features of the measured and estimated data are disparate and/or significant time delay exists between them. The spectral SE has been used several times for jet flow studies such as those of Coiffet [54] and Hall et al. [55] [56]. The spectral SE has also been recently used with HOSE by Baars et al. [57] still for an axisymmetric jet flow.

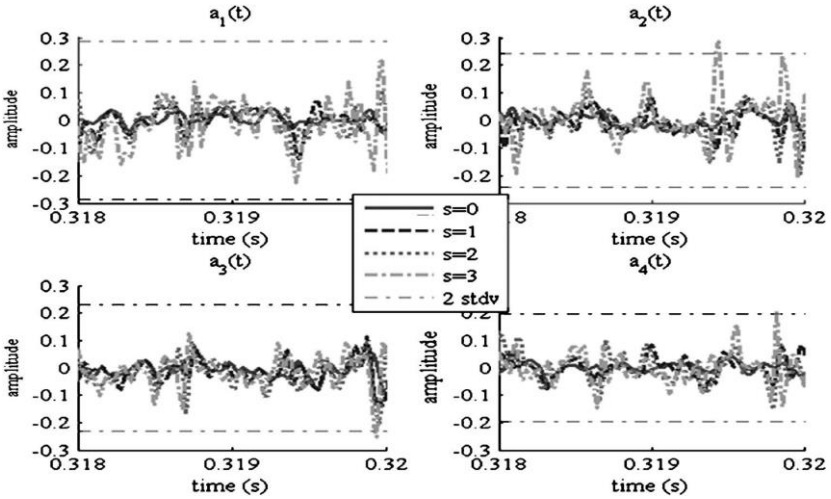


Fig. I.5: Quadratic stochastic estimation of the modal amplitude using zero to three time steps back (*s*=0–3), from experimental data in comparison to two standard deviation (2 stdv) of the modal coefficient from PIV data [39]

Alongside the spectral SE, multi-time-delay SE (MTD-SE), in the temporal domain, has also been developed. Carabello et al. [39] are among the first to use MTD-SE in 2007. In their study, using up to three time delays, they did not observe a significant effect in the linear case. In the quadratic case, the effect of adding delays was, this time, not negligible (see Fig. I.5). The MTD-SE was also used by Ukeiley et al. [58] which investigated the possibility to estimate surface pressure in a cavity from strictly past surface pressure measurements. They showed that the low frequency oscillations were correctly estimated. MTD-SE was then precisely detailed by Durgesh et al. [20] in 2010. They compared the use of MTD-LSE-POD and LSE-POD in the near wake flow of a bluff body and showed improvements in the accuracy of the reconstruction using the MTD-LSE-POD method. At last, Sicot et al. [59] developed a Spatio-Temporal POD (ST-POD) to decompose the pressure measurements, which are then used as unconditional events for estimating the velocity field by LSE. They called their technique Spatio-Temporal LSE (ST-LSE) and showed that phased average fields were remarkably well estimated by it (see Fig. I.6).

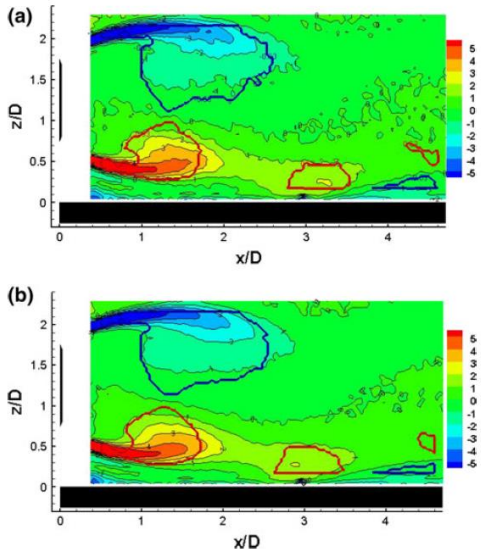


Fig. I.6: Phase-averaged velocity fields (a) and phase-averaged of the ST-LSE velocity fields for $160° < \varphi_{a1} - \varphi_{a2} < 180°$ [59]

2. Stochastic Estimation applications to real time prediction of turbulent flows

As previously described, SE was initially introduced, in the field of turbulence, as a tool to study turbulent flows through the detection of coherent structures. In this area SE has proven to be an important tool that could be used in tandem with other methods of structure identification [26] [43]. The possibility to use SE in order to reconstruct or predict the time evolution of a flow was hinted by Cole et al. in 1992 [7], who used the SE to reconstruct the temporal organization of large scale structures in a jet mixing layer. They compared the LSE estimated field with original ones, but the quality of the estimation was only quantified in term of percent of TKE estimated. In 1999, Ewing et al. [53] also use Spectral LSE and POD to study their ability to estimate the dynamic of large structures in an axisymmetric shear layer. In particular, the accuracy of the estimation is evaluated using the normalized mean square error (NMSE, see definition in Chapter II.1.4) between the estimated variable and the true one. However, in 2004, Taylor et al. [12] are the first to clearly identify and harness the possibility to predict in real-time turbulent flows using SE. Applying LSE-POD to the flow between a backward facing ramp (for Reynolds number, based on the ramp height, ranging from 2×10^4 to 5×10^4) and an adjustable flap experiments, they investigate the ability of the method to accurately estimate the time evolution of POD coefficients from wall pressure measurements. The final goal of their work was the development of active feedback control in separated flows. To fulfill this objective, they first generate a POD-based low dimensional description of the flow. They showed that in all their test cases, 4 POD modes were enough to retain more than 98% of the TKE. Then, they investigated the use of LSE to estimate the POD coefficients associated to these first 4 POD modes. Velocity information was then reconstructed for the estimated POD coefficients. They demonstrated the potential of LSE-POD to provide velocity information from wall pressure measurements, which is a requirement for active feedback flow control systems. However, in their work, the quality of the estimation is only assessed qualitatively by comparing velocity profiles (see Fig. I.7).

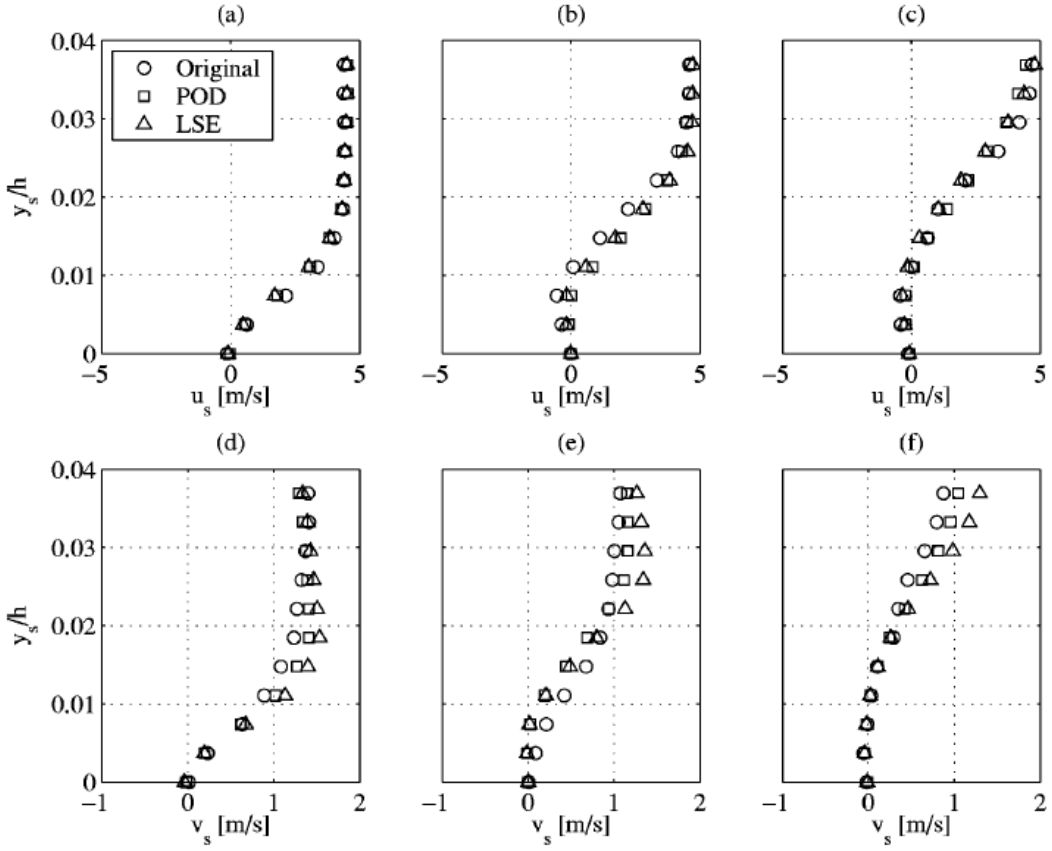


Fig. I.7: Comparison of the wall normal velocity profiles on the ramped surface, $u_s(y_s/h)$ and $v_s(y_s/h)$, at (a, d) $x_s/h = 0.34$, (b, e) $x_s/h = 0.67$, and (c, f) $x_s/h = 1.01$ with a flap angle of 21 deg and $\mathrm{Re}_h = 2 \times 10^4$ [12]

Following their work, SE is then seen as a possible tool to build Reduced-Order-Model (ROM) of the flow, or to provide real-time estimation of the flow state, in order to be used in flow control. In 2005, Rowley et al. [21] compared the performances of the LSE with a dynamic estimator, based on POD-Galerkin model, for the control of a cavity flow. They concluded that the dynamic estimator was less sensitive to measurements noise and required fewer sensors (see Fig. I.8) than the LSE. Debiasi et al. [60] then studied the application of SE to update the POD coefficients used in a POD-ROM, based on POD-Galerkin projection, in order to apply closed-loop flow control in a cavity. They compared the LSE-POD, QSE-POD, as well as their multi-time counterpart. They stated that the QSE-POD and MTD-SE were preferable than the LSE-POD. In addition, they highlighted a strong influence of the number and placement of the pressure sensors used. In continuation of this work, Carabello et al. [39] utilized LSE-POD and QSE-POD to update, in real-time, the model variable of their POD-ROM, from wall pressure measurements. The POD-ROM served to implement a feedback control for a subsonic cavity flows. Murray et al. [36] applied the QSE-POD as a first step toward producing a time-resolved ROM for a resonating subsonic cavity flow. The QSE-POD is used to produce estimation of both velocity and density in the cavity from wall pressure measurements. Their results demonstrated the ability of the QSE-POD to estimate the structure of the cavity flow field. At the same time, Pinier et al. [61] successfully implemented a proportional feedback loop using LSE-POD from wall pressure measurements to predict the state of the flow. The closed feedback loop is used to delay the separation of the flow over the NACA-4412 airfoil. They showed that the real-time prediction of the first POD coefficient provided the relevant amplitude and frequency information for driving the actuators. Stanov et al. [62] investigated experimentally the effectiveness of LSE-POD, from surface mounted hot-film sensors, to estimate a D-shaped cylinder wake, the final goal being the control of such a flow. In particular, they used a heuristic

method for placing the sensors based on the spatial content of the velocity POD modes (see Cohen et al. [63]). The quality of the estimation was evaluated with the root mean square error. They showed that this error was under 35% for the first two POD modes, which they stated as being acceptable for moderately robust control. In addition, they evaluated the estimation quality outside the training set and showed that the accuracy was deteriorated by a poor convergence of the data in the training set.

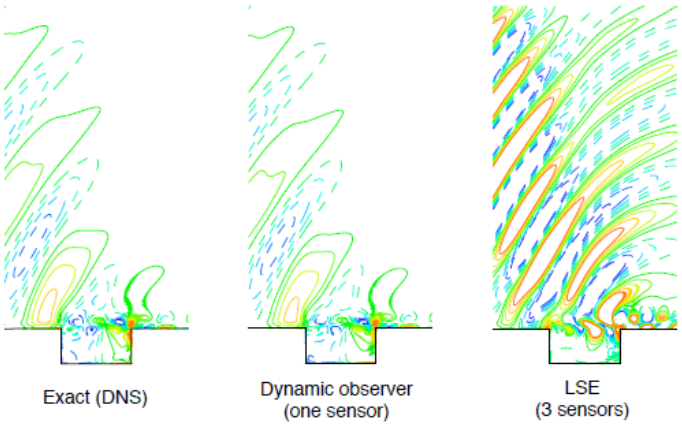


Fig. I.8: Instantaneous contours of dilatation from the exact simulation (left), and estimate from dynamic observer (center) and LSE (right), using noisy pressure signals [21]

Durgesh et al. [20] compared the use of the LSE-POD and MTD-LSE-POD to capture the near wake dynamics downstream of a bluff body (from experimental data). They showed that the inclusion of time delays was necessary to produce an accurate wake dynamic. In particular, only MTD-LSE-POD was able to recover the phase between the first two POD modes. Moreover, the choice of time delays was investigated. They observed that using delays chosen at the peaks of correlation between the POD coefficients and the pressure measurements lead to less accurate results than taking several delays around those peaks. The time window in which delays are chosen is showed to impact the estimation suggesting an optimal size of the window (see Fig. I.9). At last, if in their work Durgesh et al. used past and future time delays, they suggested that the MTD-LSE-POD would be suited for real-time applications using only past events. More precisely, they concluded that the technique should be able to capture the dynamics of any turbulent flows that has significant energy in a few POD modes. In 2010, Nguyen et al. [22] experimentally test several methods in order to predict the turbulent flow state from a limited number of real-time measurements. In their work, LSE-POD, QSE-POD, Principal Component Regression-POD (PCR-POD) and Kernel Ridge Regression-POD (KRR-POD), using single-time and multi-time-delay, are applied on a backward-facing step flow using wall shear gradient. What they call the PCR-POD can be viewed as a LSE using POD coefficients (not all of them) of the sensors as conditional events and estimating the POD coefficients of the velocity field. By using not using the highest POD modes of the sensors, it is possible to improve the conditioning of the matrix of the conditional events. The KRR-POD that they described used Radial Basis Function (RBF) network to estimate the velocity field POD coefficients, in particular they employed multi-quadratic functions. (More details on RBF are given in Chapter VI.3.5). Performances of the single-time techniques are found to be nearly identical. The MTD version of the PCR-POD and KRR-POD were found to be more accurate than the MTD-LSE-POD but only in the near wall region, not in the full domain. Lately Lasagna et al. [64] used MTD-LSE in a cavity with curvilinear geometry. They demonstrated that the MTD-LSE outperforms the LSE and QSE with significant improvements in accuracy of the estimate. The improvements were found to be strongly dependent on the number of past events used in the MTD-LSE. In addition, they also stressed the low computational effort and low number of sensors required by the method, making it particularly attractive for flow control system.

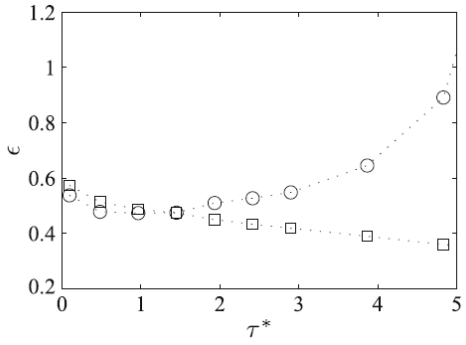


Fig. I.9: The combined error values: open square ε values for the estimation in the set of data used to determine the MTD-LSE-POD coefficients, and open circle ε values for the estimation in another set of data [20]

Recently, performances of LSE, QSE to produce accurate real-time flow state predictions have been compared with the ones of Kalman Filter (KF), Extended KF (EKF) and Linear Time-Invariant Filter (LTIF) by Sinha et al. [19] (details on Kalman Filter are provided in Chapter VI.2). In addition to serve as a direct estimator of the flow state, Sinha et al. provided another application of the LSE. They used LSE to form the linear observation operator necessary for the KF. In such application, the wall pressure measurements are reconstructed from the POD coefficients. In their work, the KF was found to have similar or better accuracy than the QSE. Both QSE and KF were shown to outperform the LSE. The EKF was however outperformed by the KF. They hinted that the uncertainty covariances were maybe improper due to an overfitting of the training data set. In 2012, Tu et al. [65] proposed to use a Kalman smoother to estimate, from non-time-resolved Particle Image Velocimetry (PIV) measurements and time-resolved velocity point sensors, time-resolved velocity fields. The main goal is thus to enrich the experimental data base. The MTD-LSE-POD (using past and future events) is first used to provide time-resolved estimates of the flow field. These estimates are then used to identify a linear dynamic model which is utilized in the Kalman smoother. They showed that the final Kalman smother estimates are more robust to noise and more accurate than the initial MTD-LSE-POD estimates (see Fig. I.10 and Fig. I.11). As Durgesh et al., they highlight the existence of an optimal window size in which delays should be chosen for the MTD-LSE-POD. In addition, Tu et al. also studied the performance of the KF and the MTD-LSE-POD using only past information. Once more the KF gave more accurate estimated velocity field.

Eventually, among the most recent works on SE applied to flow prediction are those of Clark et al., Hosseini et al. and Lasagna et al. In 2014, Clark et al. [23] proposed a study of SE from the point of view of linear least squares estimation (or ordinary least square regression). Using the linear least squares estimation formalism, they proposed a framework to evaluate the performance of SE for model construction. They applied their framework for the estimation of the wake flow behind a blunt trailing edge using MTD-LSE-POD with wall pressure measurements. Hosseini et al. [66] recently applied the Extended POD (EPOD), combined with a particular decomposition of the flow field, to the estimation of the wake of a low aspect pyramid. The flow is first decomposed into a slow-drift mode and anti-symmetric/symmetric modes. Then EPOD with multi-time is used to reconstruct the flow. The method is shown to improve the estimation. At last, Lasagna et al. [67] [68] developed a nonlinear MTD-SE. In their work, an Artificial Neural Network is used to determine the coefficients of the nonlinear MTD-SE. They compared the classical MTD-LSE with the nonlinear MTD-SE on two flows: a cavity flow and a turbulent channel flow. The inclusion of nonlinear term is found to be unnecessary in the cavity case. On the other hand, the nonlinear MTD-SE lead to a better estimation in the buffer layer of the boundary layer of the turbulent channel flow.

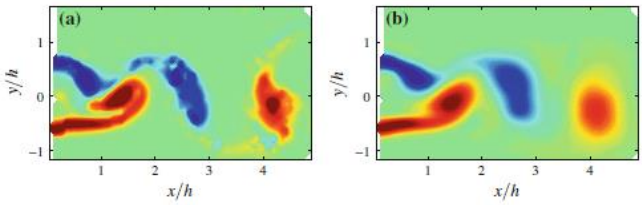


Fig. I.10: Comparison of spanwise vorticity. (a) True PIV snapshots; (b) Projection onto a seven-mode POD basis [65]

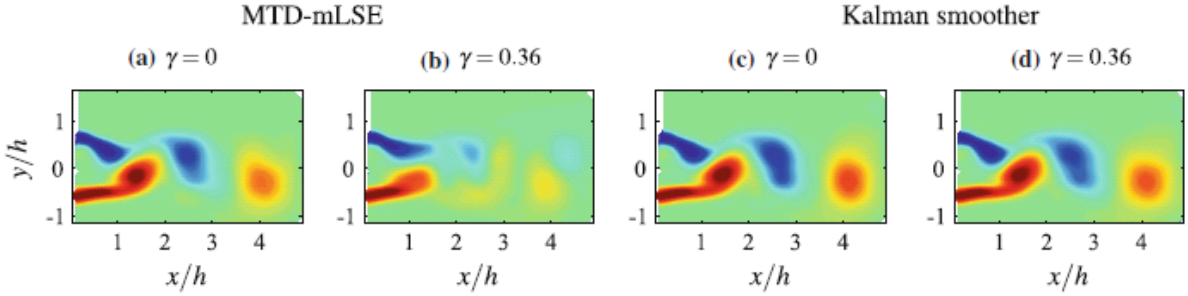


Fig. I.11: Comparison of estimated spanwise vorticity fields with and without the addition of Gaussian noise to the probe signal. (a) MTD-LSE-POD with γ = 0; (b) MTD-LSE-POD with γ = 0.36; (c) Kalman smoother with γ = 0; (d) Kalman smoother with γ = 0.36 [65]

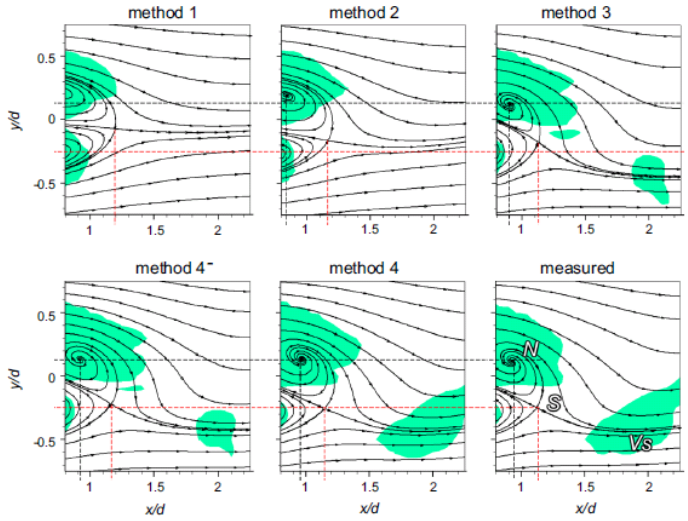


Fig. I.12: Sectional streamlines and the vortex core regions identified as $\lambda_2 \leq -0.05$ for measured and estimated velocity vector fields [66]

3. Analysis and conclusions

Since its introduction in the field of turbulence by Adrian, Stochastic Estimation (SE) has been extensively and successfully used for a large variety of turbulent flows, in order to extract coherent structures and better understand the dynamic of these flows. In Annex A, Table A.1, a

non-exhaustive summary of the studies on SE is proposed. SE has been utilized in isotropic turbulence, shear layer, axisymmetric jet, pipe flow, cavity flow, backward facing step (and ramp) flow, wake flow, thus demonstrating its attractiveness for the study of turbulent flow physics. Most often, the SE has been applied to study the flow physics through structures identification. But SE has also been applied to generate wall stress model [34] or inlet conditions for numerical simulations [48]. It has also been used as a possible tool to correct PIV database [69] or to enrich PIV database [65]. At last, SE has been applied in closed loop flow control [12] [61].

Furthermore, several versions of the SE have been developed and used. First, the non-linear SE and in particular the Quadratic SE (QSE) have been investigated. Then, two different couplings of the SE with Proper Orthogonal Decomposition (POD) have been proposed. These couplings allow forming low dimensional models of the flow dynamics. Moreover, they also reduce the computational cost. At last, Multi-Time-Delays-SE (MTD-SE) and spectral SE methods have been developed and their ability to improve the accuracy of the estimate has been proven. In addition to these extensions, the SE has also been used with a large range of conditional events. If initially inflow velocity measurements where most commonly used, it rapidly became usual to use other quantities, such as pressure, and to use nonintrusive measurements, such as surface measurements. Thus this literature overview highlights the wide variety of use of the SE.

The application of SE to the real-time prediction of turbulent flows is however more recent. Therefore, few studies strictly aimed at the investigation of the accuracy of instantaneous reconstructed flow field. Most of works in this direction concern the application of such prediction in feedback loop for flow control. In this area, most of the recent studies favored the use of MTD-SE. As stated by Sinha et al. [19] the non-recursive nature and the necessity to compute the temporal Fourier transform restrict, however, the spectral SE to offline applications. The results found in the literature showed the ability of the SE, and more particularly of the MTD-LSE-POD, to provide accurate flow prediction that can be used in flow control.

However, the applications of SE to real-time prediction of turbulent flows are limited to relatively simple turbulent flows. Taylor et al. showed results for a backward facing ramp at a Reynolds number based on the ramp height (Re_h) of 2×10^4 [12]. Rowley et al. used a cavity flow at $Re_d = 395$ (based on cavity depth d) [21], Nguyen et al. a backward facing step at $Re_h = 2.8\times10^3$ [22], Durgesh et al. the wake downstream a bluff body at $Re_h = 2\times10^4$ [20] and Tu et al. a similar flow at $Re_h = 3.6\times10^3$. Also Pinier et al. showed results for the detached flow over a NACA-4412 airfoil at $Re_c = 1.35\times10^5$ (based on the chord), but they stated their confidence that LSE-POD would be efficient at higher Reynolds number.

Furthermore, no precise methodology arises from the literature to assess the SE accuracy. Stalnov et al. [62] stressed out some questions that need to be answered concerning SE or any flow prediction methods:

- “How many sensors are required?”
- “Where should the sensors be?”
- “What are the criteria for judging an effective sensors configuration?”

For works in which the goal is the study of the flow, the accuracy of SE to produce instantaneous velocity field estimate is generally not assessed and not precisely quantified. But, for instance, Cole et al. [7] quantified the reconstructed velocity fields in term of energy captured by the estimates and Bonnet et al. [17] compared the estimated RMS velocity field to the original one. The comparison of TKE estimation is also used by Murray et al. [11] [36] to evaluate the accuracy of the estimation. If it is interesting to look at these integral characteristics of the flow, such evaluation of the SE gives no information about the quality of the instantaneous estimates.

In their work Stalnov et al. answered some of the questions. They proposed a heuristic method to choose the sensor locations and they evaluate the accuracy by means of the root-mean-square error. Recently Nguyen et al. [22] applied a more precise evaluation method to assess the accuracy of the several techniques they compared. First, they clearly distinguished the reconstruction ability of a technique from its prediction ability. The reconstruction ability corresponds to the ability to estimate data that have been used to train the technique. The prediction ability corresponds to the ability to estimate data that were not used to train the technique. Therefore, in real-time applications of the SE, one is interested in its prediction ability. Consequently, Nguyen et al. separate their database into two separate sets of data. One is used to train the estimation techniques, the other to assess their accuracy. At last, they proposed to use the mean-square error between the true and estimated POD coefficients as well as a correlation coefficient to quantify the accuracy of the estimate. Using this evaluation approach, they concluded that the QSE-POD had the best reconstruction ability but the worse prediction ability. Thus they highlighted that the QSE-POD suffered from overfitting in their case. The use of two separate data sets to assess the accuracy of SE has then been used by Sinha et al. [19], Durgesh et al. [20] and Tu et al. [65]. Durgesh et al., as well as Tu et al., showed then that the reconstruction ability of the MTD-SE was always improved by the addition of new time delay whereas the prediction ability deteriorated when too many delays were used (see Fig. I.9). Because such precise evaluation was not clearly performed in other studies comparing LSE and QSE, it is, therefore, difficult to judge if the conclusions that were drawn concerned the reconstruction ability or the prediction ability of the QSE. Only in the work of Sinha is the QSE prediction accuracy clearly better than the LSE one. Another quantification criterion than the ones used by Nguyen et al., has also been proposed. Sinha et al., Durgesh et al. and Tu et al. used an error metric that is the ratio of combined residual variance to the total variance of the POD basis used. It is calculated by:

$$e_{POD} = \frac{\langle \sum_{i=1}^{N_{POD}} (\tilde{a}_i(t) - a_i(t))^2 \rangle}{\langle \sum_{i=1}^{N_{POD}} a_i(t)^2 \rangle}$$

More recently this error metric has also been utilized by Clark et al. [23] and Hosseini et al. [66]. A similar error metric directly applied to the velocity field has also been employed by Hosseini et al. and Lasagna et al. [68]. Clark et al. also proposed to use the Variance Inflation Factor as a criterion to evaluate the degree of collinearities between the measurements and thus the risk of overfitting.

At last, from this literature overview, it appears that the SE has rarely been used in the goal to provide in real time precise aerodynamic information. If the use of SE in flow control is close to such application of the SE, it remains that the level of information required in flow control is lower.

Chapter II. Methods and databases

The goal of this chapter is first to present the mathematical framework of the Stochastic Estimation and its extensions. LSE, QSE, LSE-POD, MTD-LSE are detailed. That way, some of their characteristics can be highlighted. In addition, prerequisite to use SE are explained, especially in terms of necessary data.

From these considerations and the literature overview on SE, test cases are chosen. Indeed, the studies of SE are mainly based on the application of the methods on test cases that can be obtained experimentally or numerically. Thus the available experimental apparatuses are first described and the constraints they imposed are stressed out. Then the experimental databases are described thoroughly, especially in terms of convergence of the statistical moments.

At last, the numerical databases are also presented in a similar manner as the experimental ones.

1. Stochastic Estimation mathematical framework

1.1. Single-Time Stochastic Estimation

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

$$\tilde{u}(\boldsymbol{x},t) = \langle u(\boldsymbol{x},t)|\mathbf{E}\rangle \tag{II.1}$$

$\tilde{u}$ is the expected value of u under the condition **E**. If the quantity *u* is assumed to be continuous and of zero mean, then its Taylor expansion around the average value of **E** is:

$$\tilde{u}(\boldsymbol{x},t) = \sum_{i=1}^{N} \mathrm{A}_{\boldsymbol{x}}(y_i)\,\mathbf{E}(y_i,t) + \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{B}_{\boldsymbol{x}}\left(y_i,y_j\right)\mathbf{E}(y_i,t)\mathbf{E}(y_j,t) + \cdots \tag{II.2}$$

N is the number of y_i points where **E** is known (for instance the number of sensors used).

Keeping only first order terms in the Taylor expansion leads to the linear estimation of the expected value $\tilde{u}$. Thus it is called the LSE. Keeping first and second order terms leads to the QSE. At last, keeping higher order terms is usually referenced as HOSE. In any cases, A_x and B_x (or higher) coefficients are calculated through the minimization of the mean square error between the SE $\tilde{u}$ and the true value *u* on a training set of data:

$$e(\boldsymbol{x}) = \langle(\tilde{u}(\boldsymbol{x},t) - u(\boldsymbol{x},t))^2\rangle \tag{II.3}$$

where $\langle . \rangle$ denotes the ensemble average. In the discrete case, the training set is a collection of statistically independent realizations of both $u(\boldsymbol{x},t)$ and **E**(*t*). Let us assume that the training set contains *M* realizations denoted by t_i for $i \in [\![1,M]\!]$, then the ensemble average is:

$$\langle u(\boldsymbol{x},t)\rangle = \frac{1}{M}\sum_{i=1}^{M} u(\boldsymbol{x},t_i) \tag{II.4}$$

In the linear case, the minimization problem requires to solve the *N* equations:

$$\frac{\partial e(\boldsymbol{x})}{\partial \mathrm{A}_{\boldsymbol{x}}(y_i)} = 0, \forall i \in [\![1,N]\!]$$

$$\Leftrightarrow \frac{\partial\left(\langle\left(\sum_{j=1}^{N}\mathrm{A}_{\boldsymbol{x}}(y_j)\mathbf{E}(y_j,t)\right)^2\rangle - 2\langle\sum_{j=1}^{N}\mathrm{A}_{\boldsymbol{x}}(y_j)\mathbf{E}(y_j,t)u(\boldsymbol{x},t)\rangle + \langle u(\boldsymbol{x},t)^2\rangle\right)}{\partial \mathrm{A}_{\boldsymbol{x}}(y_i)} = 0,$$

$$\forall i \in [\![1,N]\!] \tag{II.5}$$

$$\Leftrightarrow \langle 2\mathbf{E}(y_i,t)\sum_{j=1}^{N}\mathrm{A}_{\boldsymbol{x}}(y_j)\mathbf{E}(y_j,t)\rangle - 2\langle\mathbf{E}(y_i,t)u(\boldsymbol{x},t)\rangle = 0, \forall i \in [\![1,N]\!]$$

$$\Leftrightarrow \langle u(\boldsymbol{x},t)\mathbf{E}(y_i,t)\rangle = \sum_{j=1}^{N}\mathrm{A}_{\boldsymbol{x}}(y_j)\langle\mathbf{E}(y_j,t)\mathbf{E}(y_i,t)\rangle, \forall i \in [\![1,N]\!]$$

Which leads to the resolution of the following matrix system for any **x** at which one wants to estimate the scalar velocity field $u(\boldsymbol{x},t)$:

$$\begin{pmatrix}\mathrm{A}_{\boldsymbol{x}}(y_1)\\ \vdots \\ \mathrm{A}_{\boldsymbol{x}}(y_N)\end{pmatrix} = \begin{pmatrix}\langle\mathbf{E}(y_1,t)\mathbf{E}(y_1,t)\rangle & \dots & \langle\mathbf{E}(y_1,t)\mathbf{E}(y_N,t)\rangle \\ \vdots & \ddots & \vdots \\ \langle\mathbf{E}(y_N,t)\mathbf{E}(y_1,t)\rangle & \dots & \langle\mathbf{E}(y_N,t)\mathbf{E}(y_N,t)\rangle\end{pmatrix}^{-1}\begin{pmatrix}\langle u(\boldsymbol{x},t)\mathbf{E}(y_1,t)\rangle \\ \vdots \\ \langle u(\boldsymbol{x},t)\mathbf{E}(y_N,t)\rangle\end{pmatrix} \tag{II.6}$$

The problem of finding the coefficients $\mathrm{A}_{\boldsymbol{x}}(y_i)$ is, in fact, equivalent to the overdetermined system:

$$\mathrm{EA} = \mathrm{U} \tag{II.7}$$

with:

$$\mathrm{E} = \begin{pmatrix} \mathbf{E}(y_1, t_1) & \cdots & \mathbf{E}(y_N, t_1) \\ \vdots & \ddots & \vdots \\ \mathbf{E}(y_1, t_M) & \cdots & \mathbf{E}(y_N, t_M) \end{pmatrix}, \mathrm{A} = \begin{pmatrix} \mathrm{A}_{\boldsymbol{x}}(y_1) \\ \vdots \\ \mathrm{A}_{\boldsymbol{x}}(y_N) \end{pmatrix} \text{ and } \mathrm{U} = \begin{pmatrix} u(\boldsymbol{x}, t_1) \\ \vdots \\ u(\boldsymbol{x}, t_M) \end{pmatrix}$$

And the Ordinary Least Square solution of this system of equations is $\mathrm{A} = (\mathrm{EE}^{\mathrm{T}})^{-1}\mathrm{E}^{\mathrm{T}}\mathrm{U}$ wich is equation (II.6).

In a completely similar manner, the overdetermined system to solve, in the quadratic case, is still equation (II.7) but with the following matrices:

$$\mathrm{E} \in \mathbb{R}^{M \times \left(N + \frac{N \times (N-1)}{2}\right)}$$

$$= \begin{pmatrix} \mathbf{E}(y_1, t_1) & \cdots & \mathbf{E}(y_N, t_1) & \mathbf{E}(y_1, t_1)\mathbf{E}(y_1, t_1) & \ldots & \mathbf{E}(y_2, t_1)\mathbf{E}(y_2, t_1) & \ldots & \mathbf{E}(y_{\mathrm{N}}, t_1)\mathbf{E}(y_{\mathrm{N}}, t_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{E}(y_1, t_M) & \cdots & \mathbf{E}(y_N, t_M)\mathbf{E}(y_1, t_M)\mathbf{E}(y_1, t_M) & \ldots & \mathbf{E}(y_2, t_M)\mathbf{E}(y_2, t_M) & \ldots & \mathbf{E}(y_{\mathrm{N}}, t_M)\mathbf{E}(y_{\mathrm{N}}, t_M) \end{pmatrix}$$

$$\mathrm{A} = \begin{pmatrix} \mathrm{A}_{\boldsymbol{x}}(y_1) \\ \vdots \\ \mathrm{A}_{\boldsymbol{x}}(y_N) \\ \mathrm{B}_{\boldsymbol{x}}(y_1, y_1) \\ \vdots \\ \mathrm{B}_{\boldsymbol{x}}(y_N, y_N) \end{pmatrix}$$

U remains unchanged. A is thus solution of the matrix system:

$$\mathrm{A} = \mathrm{C}^{-1}\mathrm{B} \tag{II.8}$$

With:

$$\mathrm{A} = \begin{pmatrix} \mathrm{A}_{\boldsymbol{x}}(y_1) \\ \vdots \\ \mathrm{A}_{\boldsymbol{x}}(y_N) \\ \mathrm{B}_{\boldsymbol{x}}(y_1, y_1) \\ \vdots \\ \mathrm{B}_{\boldsymbol{x}}(y_N, y_N) \end{pmatrix}, \mathrm{C} = \mathrm{EE}^{\mathrm{T}} = \begin{pmatrix} \mathrm{C}_{11} & (\mathrm{C}_{12})^{\mathrm{T}} \\ \mathrm{C}_{12} & \mathrm{C}_{22} \end{pmatrix}, \mathrm{B} = \mathrm{E}^{\mathrm{T}}\mathrm{U} = \begin{pmatrix} \langle u(\boldsymbol{x}, t)\mathbf{E}(y_1, t) \rangle \\ \vdots \\ \langle u(\boldsymbol{x}, t)\mathbf{E}(y_N, t) \rangle \\ \langle u(\boldsymbol{x}, t)\mathbf{E}(y_1, t)\mathbf{E}(y_1, t) \rangle \\ \vdots \\ \langle u(\boldsymbol{x}, t)\mathbf{E}(y_N, t)\mathbf{E}(y_N, t) \rangle \end{pmatrix}$$

$$\mathrm{C}_{11} = \begin{pmatrix} \langle \mathbf{E}(y_1)\mathbf{E}(y_1) \rangle & \ldots & \langle \mathbf{E}(y_1)\mathbf{E}(y_N) \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{E}(y_N)\mathbf{E}(y_1) \rangle & \ldots & \langle \mathbf{E}(y_N)\mathbf{E}(y_N) \rangle \end{pmatrix}$$

$$\mathrm{C}_{12} = \begin{pmatrix} \langle \mathbf{E}(y_1)\mathbf{E}(y_1)\mathbf{E}(y_1) \rangle & \ldots & \langle \mathbf{E}(y_1)\mathbf{E}(y_1)\mathbf{E}(y_N) \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{E}(y_N)\mathbf{E}(y_N)\mathbf{E}(y_1) \rangle & \ldots & \langle \mathbf{E}(y_N)\mathbf{E}(y_N)\mathbf{E}(y_N) \rangle \end{pmatrix}$$

$$\mathrm{C}_{22} = \begin{pmatrix} \langle \mathbf{E}(y_1)\mathbf{E}(y_1)\mathbf{E}(y_1)\mathbf{E}(y_1) \rangle & \cdots & \langle \mathbf{E}(y_1)\mathbf{E}(y_1)\mathbf{E}(y_N)\mathbf{E}(y_N) \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{E}(y_N)\mathbf{E}(y_N)\mathbf{E}(y_1)\mathbf{E}(y_1) \rangle & \cdots & \langle \mathbf{E}(y_N)\mathbf{E}(y_N)\mathbf{E}(y_N)\mathbf{E}(y_N) \rangle \end{pmatrix}$$

The number of B_x coefficients is equal to $\frac{N\times(N-1)}{2}$. Indeed, since $\langle \mathbf{E}(y_i)\mathbf{E}(y_j)\rangle = \langle \mathbf{E}(y_j)\mathbf{E}(y_i)\rangle$, then one can consider only the $B_{\boldsymbol{x}}(y_i, y_j)\ \forall (i,j) \in [\![1, N]\!]^2$ and $i \leq j$.

From the way the LSE is introduced here, one can clearly see its relationship to the Ordinary Least Square regression. This point has been precisely discussed recently by Clark et al. [23]. The LSE requires the covariance matrix between the conditional events, as well as the covariance matrix between the conditional events and the quantity to estimate. In QSE, statistical moments up to order 4 are required. A prerequisite for a correct and meaningful estimation outside of the training set, is thus to obtain converged moments from the data in the training set (see the discussion about over-fitting in Chapter II. 1.4). The uncertainties of the statistical moments can be assessed using formula provided by Benedict et al. [70] (see Chapter II.3. and Chapter II.4.).

Once the SE coefficients ($A_{\boldsymbol{x}}$, $B_{\boldsymbol{x}}$ or higher) have been computed on a 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 $\boldsymbol{u}$**, whereas the estimation of u in the training set is referred as **the reconstruction of $\boldsymbol{u}$**.

The SE owns some properties. In the case where the quantity to estimate is a vector field $\boldsymbol{u}$ satisfying the continuity equation, then the SE estimated vector field also satisfies the continuity equation [27]. Simply put, in the LSE case, from equation (II.5):

$$\langle \mathrm{div}\left(\boldsymbol{u}(\boldsymbol{x},t)\right)\mathbf{E}(y_i,t)\rangle = \sum_{j=1}^{N} \mathrm{div}\left(A_{\boldsymbol{x}}(y_j)\right)\langle \mathbf{E}(y_j,t)\mathbf{E}(y_i,t)\rangle\,, \forall i \in [\![1,N]\!]$$

Thus, if the vector field satisfies the continuity equation, then the following system of equations stands:

$$\begin{pmatrix} \langle \mathbf{E}(y_1,t)\mathbf{E}(y_1,t)\rangle & \dots & \langle \mathbf{E}(y_1,t)\mathbf{E}(y_N,t)\rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{E}(y_N,t)\mathbf{E}(y_1,t)\rangle & \dots & \langle \mathbf{E}(y_N,t)\mathbf{E}(y_N,t)\rangle \end{pmatrix} \begin{pmatrix} \mathrm{div}\left(A_{\boldsymbol{x}}(y_1)\right) \\ \vdots \\ \mathrm{div}\left(A_{\boldsymbol{x}}(y_N)\right) \end{pmatrix} = \begin{pmatrix} \langle \mathrm{div}\left(\boldsymbol{u}(\boldsymbol{x},t)\right)\mathbf{E}(y_1,t)\rangle \\ \vdots \\ \langle \mathrm{div}\left(\boldsymbol{u}(\boldsymbol{x},t)\right)\mathbf{E}(y_N,t)\rangle \end{pmatrix} = 0$$

In addition, if the system (II.6) has a solution, then:

$$\det \begin{pmatrix} \langle \mathbf{E}(y_1,t)\mathbf{E}(y_1,t)\rangle & \dots & \langle \mathbf{E}(y_1,t)\mathbf{E}(y_N,t)\rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{E}(y_N,t)\mathbf{E}(y_1,t)\rangle & \dots & \langle \mathbf{E}(y_N,t)\mathbf{E}(y_N,t)\rangle \end{pmatrix} \neq 0$$

Therefore $\mathrm{div}\left(A_{\boldsymbol{x}}(y_i)\right) = 0, \forall i \in [\![1,N]\!]$ and $\mathrm{div}\left(\tilde{u}(\boldsymbol{x},t)\right) = \sum_{i=1}^{N} \mathrm{div}\left(A_{\boldsymbol{x}}(y_i)\right)\mathrm{E}(y_i,t) = 0$.

Another property of the SE is that, if the quantity to estimate and the conditional event **E** are of similar nature, then the estimation at a position where **E** is known is exact. Moreover, the SE is inhomogeneous in the mean that it is possible to estimate a quantity from conditional events of different physical nature. As one can see from equation (II.6), LSE uses the covariance between the quantity to estimate and the conditional event. Thus a quantity not correlated to the conditional event cannot be estimated. The choice of the conditional event (how many and where) is therefore of crucial importance.

In 1989, Guezennec proposed a STD-SE [15]. This version of the SE consists in considering phase-shifted in time conditional events. The equation (II.2) becomes:

$$\tilde{u}(\boldsymbol{x},t)=\sum_{i=1}^{N}\mathrm{A}_{\boldsymbol{x}}(y_i)\,\mathbf{E}(y_i,t-\tau)+\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{B}_{\boldsymbol{x}}\left(y_i,y_j\right)\mathbf{E}(y_i,t-\tau)\mathbf{E}(y_j,t-\tau)+\cdots \qquad \text{(II.9)}$$

with τ a delay. The systems (II.6) and (II.8) are modified accordingly.

1.2. Multi-Time-Delay Stochastic Estimation

Extending the SE method to multi-time simply consists in using several realizations of the conditional event **E**, not only in space, but also in time. The estimation of the random quantity $u(\boldsymbol{x},t)$ then becomes:

$$\begin{aligned}\tilde{u}(\boldsymbol{x},t)=&\sum_{i=1}^{N}\sum_{j=1}^{N_d}\mathrm{A}_{\boldsymbol{x}}\left(y_i,\tau_j\right)\mathbf{E}\left(y_i,t-\tau_j\right)\\&+\sum_{k=1}^{N_d}\sum_{l=1}^{N_d}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{B}_{\boldsymbol{x}}\left(y_i,y_j,\tau_k,\tau_l\right)\mathbf{E}(y_i,t-\tau_k)\mathbf{E}(y_j,t-\tau_l)+\cdots\end{aligned} \qquad \text{(II.10)}$$

where N_d is the number of delays and τ_j the delays.

The MTD-SE coefficients are calculated in a similar manner as in single-time SE. In the MTD-LSE case, the coefficients $\mathrm{A}_{\boldsymbol{x}}\left(y_i,\tau_j\right)$ are computed by solving the linear system:

$$\mathrm{A}=\mathrm{C}^{-1}\mathrm{B} \qquad \text{(II.8)}$$

with:

$$\mathrm{A}=\begin{pmatrix}\mathrm{A}_{\boldsymbol{x}}(y_1,\tau_1)\\\vdots\\\mathrm{A}_{\boldsymbol{x}}(y_N,\tau_1)\\\mathrm{A}_{\boldsymbol{x}}(y_1,\tau_2)\\\vdots\\\mathrm{A}_{\boldsymbol{x}}(y_N,\tau_{N_d})\end{pmatrix},\mathrm{B}=\begin{pmatrix}\langle u(\boldsymbol{x},t)\mathbf{E}(y_1,t-\tau_1)\rangle\\\vdots\\\langle u(\boldsymbol{x},t)\mathbf{E}(y_N,t-\tau_1)\rangle\\\langle u(\boldsymbol{x},t)\mathbf{E}(y_1,t-\tau_2)\rangle\\\vdots\\\langle u(\boldsymbol{x},t)\mathbf{E}(y_N,t-\tau_{N_d})\rangle\end{pmatrix}$$

$$\mathrm{C}=\begin{pmatrix}\langle\mathbf{E}(y_1,t-\tau_1)\mathbf{E}(y_1,t-\tau_1)\rangle&\cdots&\langle\mathbf{E}(y_1,t-\tau_1)\mathbf{E}(y_N,t-\tau_{N_d})\rangle\\\vdots&\ddots&\vdots\\\langle\mathbf{E}(y_N,t-\tau_1)\mathbf{E}(y_1,t-\tau_1)\rangle&\cdots&\langle\mathbf{E}(y_N,t-\tau_1)\mathbf{E}(y_N,t-\tau_{N_d})\rangle\\\langle\mathbf{E}(y_1,t-\tau_2)\mathbf{E}(y_1,t-\tau_1)\rangle&\cdots&\langle\mathbf{E}(y_N,t-\tau_2)\mathbf{E}(y_N,t-\tau_{N_d})\rangle\\\vdots&\ddots&\vdots\\\langle\mathbf{E}(y_N,t-\tau_{N_d})\mathbf{E}(y_1,t-\tau_1)\rangle&\cdots&\langle\mathbf{E}(y_N,t-\tau_{N_d})\mathbf{E}(y_N,t-\tau_{N_d})\rangle\end{pmatrix}$$

In this situation, matrix dimensions are: $\mathrm{A}\in\mathbf{R}^{(N\times N_d)\times 1}$, $\mathrm{B}\in\mathbf{R}^{(N\times N_d)\times 1}$, $\mathrm{C}\in\mathbf{R}^{(N\times N_d)\times(N\times N_d)}$.

LSE, QSE and MTD-LSE have been implemented using C++ language. Several ways of solving the matrix system (II.8) have been implemented. The routine dsysv, based on LU decomposition, of the LAPACK/CBLAS library can be used. QR decomposition (through the routines dgeqrf and dormqr) or SVD (through dgesvd) are also available. The resolution using SVD was found to hold slightly more accurate results and is therefore favored. At last, Tikhonov regularization can be applied with any of the three previous methods of matrix system resolution. Details on the Tikhonov regularization are provided in annex A.

1.3. Stochastic Estimation coupled with Proper Orthogonal Decomposition

1.3.1. General reminder on Proper Orthogonal Decomposition

Proper Orthogonal decomposition is a well-known technique for flow decomposition and in particular for coherent structures identification. The idea proposed by Lumley is to use an energetic criterion, more precisely, to consider that coherent structures have the largest projection on the velocity field (according to the L2-norm). This concept leads to the resolution of an eigenvalue problem which kernel is the two-point spatial correlation tensor.

In a more general context, the POD method finds deterministic functions $\boldsymbol{\Phi}$ maximizing their projection on a random quantity $\boldsymbol{u}$. Be $\boldsymbol{X} = (\boldsymbol{x},t)$ such as $\boldsymbol{X} \in D = \Omega \times \mathbf{R}$ (Ω is the spatial domain) and let us assume that $\boldsymbol{u}(\boldsymbol{X})$ is a complex vector field with real parameters and that it is square-integral. $\boldsymbol{u} \in L_2(D)$, then the POD problem is to find $\boldsymbol{\Phi}$ maximizing:

$$\frac{\langle |(\boldsymbol{u}, \boldsymbol{\Phi})|^2 \rangle}{\|\boldsymbol{\Phi}\|^2} \tag{II.11}$$

where $\langle . \rangle$ is an average operator, $|.|$ the module, $(,)$ is a dot product on $L_2(D)$ and $\|.\|^2$ the associated norm. This maximization problem leads to solving an eigenvalue problem under a Fredholm integral equation (N_c is the number of components of $\boldsymbol{u}$):

$$\sum_{j=1}^{N_c} \int_D R_{ij}\left(\boldsymbol{X}, \boldsymbol{X}'\right) \Phi_j\left(\boldsymbol{X}'\right) \mathrm{d}\boldsymbol{X}' = \lambda \Phi_i(\boldsymbol{X}) \tag{II.12}$$

According to Hilbert-Schmidt theory, D has to be bounded and the kernel has to be part of the space of square-integral functions and be Hermitian. It is important to notice that, if it exists a homogeneous or stationary direction, then the POD cannot be used. Nevertheless, satisfying the previous conditions leads to a countable infinity of solutions $\boldsymbol{\Phi}^{(n)}$ for the Fredhlom equation.

1.3.1.1. Classic method

The approach proposed by Lumley in 1967 [42] consists in considering the average operator $\langle . \rangle$ to be a temporal average. The temporal average is evaluated using stationarity and ergodicity hypothesis through an ensemble average (as defined in equation (II.4)). The **X** variable is assimilated to the spatial variable **x** and the dot product is defined by:

$$(\boldsymbol{u}, \boldsymbol{\Phi})_\Omega = \int_\Omega \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{\Phi}^*(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \sum_{i=1}^{N_c} \int_\Omega u_i(\boldsymbol{x}) \boldsymbol{\Phi}_i^*(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{II.13}$$

The associated norm is thus:

$$\|\boldsymbol{u}\|^2 = \int_\Omega \mathbf{u}(\boldsymbol{x}) \boldsymbol{u}^*(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{II.14}$$

(Therefore if $\boldsymbol{u}$ is the fluctuating velocity field, then $\|\boldsymbol{u}\|^2$ is the integrated kinetic energy). And the Fredholm equation to solve becomes:

$$\sum_{j=1}^{N_c} \int_{\Omega} R_{ij}(\boldsymbol{x}, \boldsymbol{x}') \, \Phi_j(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}' = \lambda \Phi_i(\boldsymbol{x}) \tag{II.15}$$

The kernel of this equation is the two-point spatial correlation tensor, which considering stationary conditions and ergodicity is expressed as:

$$R_{ij}(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{T} \int_{T} u_i(\boldsymbol{x}, t) u_j(\boldsymbol{x}', t) \mathrm{d}t \, , T \to \infty \tag{II.16}$$

Assuming that the conditions of the Hilbert-Schmidt theory are satisfied, the Fredholm equation has a countable infinity of solutions (called POD modes) that can be chosen orthonormal:

$$\left(\boldsymbol{\Phi}^{(m)}, \boldsymbol{\Phi}^{(n)}\right)_{\Omega} = \delta_{mn}$$

And $u_i(\boldsymbol{x},t)$ can be decomposed as:

$$u_i(\boldsymbol{x}, t) = \sum_{n=1}^{\infty} a^{(n)}(t) \Phi_i^{(n)}(\boldsymbol{x}) \tag{II.17}$$

where $a^{(n)}$ are the projection coefficients (also called instantaneous coefficients or POD coefficients), which are defined by:

$$a^{(n)}(t) = \left(\boldsymbol{u}, \Phi^{(n)}\right)_{\Omega} = \sum_{i=1}^{N_c} \int_{\Omega} u_i(\boldsymbol{x}, t) \Phi_i^{(n)}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{II.18}$$

In addition, the POD coefficients are not correlated:

$$\langle a^{(m)} a^{(n)} \rangle = \delta_{mn} \sqrt{\lambda^{(m)} \lambda^{(n)}}$$

The eigenvalues $\lambda^{(n)}$ are thus representative of the energy contained in each mode. Since the correlations tensor is positive, they are positive. At last it is common to organize the POD modes such as:

$$\lambda^{(1)} \geq \lambda^{(2)} \geq \lambda^{(3)} \geq \cdots \geq 0$$

The POD projection on the first N_{POD} modes is thus the finite sum:

$$\hat{u}_i(\boldsymbol{x}, t) = \sum_{n=1}^{N_{POD}} a^{(n)}(t) \Phi_i^{(n)}(\boldsymbol{x}) \tag{II.19}$$

1.3.1.2. Snapshots method

The snapshots method has been proposed by Sirovich in 1987 [71]. In this approach, the average operator $\langle . \rangle$ is a spatial average defined by:

$$\langle . \rangle = \int_{\Omega} . \, \mathrm{d}\boldsymbol{x}$$

The variable $\boldsymbol{X}$ in equation (II.12) is assimilated to t. The Fredholm equation to solve is then:

$$\frac{1}{T}\int_T C(t,t')\,a_j(t')\mathrm{d}t' = \lambda a_j(t) \tag{II.20}$$

Where the spatial correlations tensor has been replaced by the temporal correlations tensor:

$$C(t,t') = \sum_{i=1}^{N_c}\int_\Omega u_i(\boldsymbol{x},t)u_i(\boldsymbol{x},t')\,\mathrm{d}\boldsymbol{x} \tag{II.21}$$

Once more this equation has a countable infinity of solutions $a^{(n)}(t)$ and the spatial modes $\boldsymbol{\Phi}^{(n)}(\boldsymbol{x})$ are obtained by projecting the random variable $\boldsymbol{u}$ on the coefficients $a^{(n)}(t)$:

$$\Phi_i^{(n)}(\boldsymbol{x}) = \int_T u_i(\boldsymbol{x},t)a^{(n)}(t)\mathrm{d}t \tag{II.22}$$

As in the classic method, it is possible to obtain orthonormal POD modes. Equations (II.17), (II.18) and (II.19) are still verified.

While the classic POD method is well suited for data which are temporally well resolved, the "snapshots" method is more adapted to data spatially well resolved. In addition, in discrete cases, if the number of temporal realizations is smaller than the spatial dimensions, then using the "snapshots" method allow to decrease the dimensions of the eigenvalue problem to solve. Both methods lead to strictly identical POD basis.

1.3.1.3. Practical implementation of the Proper Orthogonal Decomposition

Let us assume that we dispose of L temporal realizations (t_i) of the random variable $\boldsymbol{u}(\boldsymbol{x},t)$ in M discrete points of Ω. These realizations are referred to as snapshots and are stored in matrix format:

$$U = (\boldsymbol{u}(\boldsymbol{x},t_0) \quad \ldots \quad \boldsymbol{u}(\boldsymbol{x},t_L)) \in \mathbf{R}^{(M\times N_c)\times L}$$

- Classic implementation:

The classic method has been implemented in the case where the random variable u is a real scalar only. In such a case, the spatial correlation tensor is the matrix:

$$R(\boldsymbol{x},\boldsymbol{x}') = \frac{1}{T}UU^T \in \mathbf{R}^{M\times M}$$

This matrix is symmetric and positive. Thus it is diagonalizable in an orthonormal basis. The number of non-zero eigenvalues is equal to the matrix rank r. Therefore, there are $\lambda_1 \geq \ldots \geq \lambda_r \geq 0$ eigenvalues and $v_k \in \mathbf{R}^M$ ($k \in [\![1,r]\!]$) associated orthonormal eigenvectors of R. The diagonalization of R is performed using the routine dsyev of the LAPACK/BLAS library.

The POD modes are thus the eigenvectors and the coefficients POD are obtained using equation (II.18).

- Snapshots method implementation:

The snapshots method has been implemented for a real vector field. The temporal correlations tensor is:

$$C(t,t') = U^TPU \in \mathbf{R}^{L\times L}$$

where P is a symmetric positive-definite matrix. P is the matrix of the chosen dot product (.,.). By default, it is a diagonal matrix. Each element of the diagonal is the volume of the corresponding mesh cell. C is diagonalizable in an orthonormal basis. Therefore, there are $\lambda_1 \geq \ldots \geq \lambda_r \geq 0$ eigenvalues and $v_k \in \mathbf{R}^L$ ($k \in [\![1, r]\!]$) associated orthonormal eigenvectors of C. To simplify the computation and take advantage of the matrix product, we deviate from the traditional definition and the POD coefficients are defined by:

$$a^{(n)}(t) = \sqrt{N} v_n$$

And the POD modes are defined by:

$$\boldsymbol{\Phi}^{(n)} = \frac{1}{\sqrt{N}} U v_n \in \mathbf{R}^{(M \times N_c) \times 1}$$

Thus, the POD modes are only orthogonal (and not orthonormal):

$$\left(\boldsymbol{\Phi}^{(n)}, \boldsymbol{\Phi}^{(m)}\right)_\Omega = \sqrt{\lambda^{(n)} \lambda^{(m)}} \delta_{nm}$$

And:

$$\frac{1}{N} \sum_{i=1}^{N} a^{(n)}(t_i) a^{(m)}(t_i) = \delta_{nm}$$

At last, for any t_i, $i \in [\![1, L]\!]$:

$$\boldsymbol{u}(\boldsymbol{x}, t_i) = \sum_{k=1}^{r} a^{(k)}(t_i)\, \boldsymbol{\Phi}^{(k)}(\boldsymbol{x})$$

1.3.2. Complementary technique

In the complementary technique, first SE is executed to estimate a quantity. Then the estimates are projected onto the POD basis. At last, the quantity is reconstructed from the POD projection.

So the POD coefficients are obtained with:

$$a_{\text{est}}^{(n)}(t) = \left(\tilde{\boldsymbol{u}}, \boldsymbol{\Phi}^{(n)}\right)_\Omega = \sum_{i=1}^{N_c} \int_\Omega \tilde{u}_i(\boldsymbol{x}, t) \Phi_i^{(n)}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{II.23}$$

where $\tilde{\boldsymbol{u}}$ is an estimation of $\boldsymbol{u}$ using any SE methods. Then the complementary technique estimation of $\boldsymbol{u}$ is:

$$\hat{\tilde{\boldsymbol{u}}}(\boldsymbol{x}, t) = \sum_{n=1}^{N_{\text{POD}}} a_{\text{est}}^{(n)}(t) \boldsymbol{\Phi}^{(n)}(\boldsymbol{x}) \tag{II.24}$$

1.3.3. Modified Stochastic Estimation (SE-POD)

In the SE-POD, the POD coefficients of a quantity are directly estimated by SE. Then, the estimated POD coefficients are used in order to obtain an estimated POD projection of the quantity. Let us denote $\tilde{a}^{(n)}(t)$ the POD coefficients estimated by a SE method, then the SE-POD of the quantity $\boldsymbol{u}$ is:

$$\widetilde{\widetilde{\boldsymbol{u}}}(\boldsymbol{x},t)=\sum_{n=1}^{N_{\mathrm{POD}}}\tilde{a}^{(n)}(t)\,\boldsymbol{\Phi}^{(n)}(\boldsymbol{x}) \tag{II.25}$$

1.4. Errors and score parameters

In the literature on SE, few discussions addressed precisely the problem of assessing the accuracy of the estimation. In recent papers, using SE-POD, the ratio of combined residual variance to the total variance of the POD basis is more and more commonly used (see Chapter I.2). It is calculated by:

$$e_{\mathrm{POD}}=\frac{\langle\sum_{i=1}^{N_{\mathrm{POD}}}(\tilde{a}_i(t)-a_i(t))^2\rangle}{\langle\sum_{i=1}^{N_{\mathrm{POD}}}a_i(t)^2\rangle} \tag{II.26}$$

In what follow, the determination coefficient R^2 will be used to assess the quality of the estimation and is defined by $R^2_{\mathrm{POD}}=1-e_{\mathrm{POD}}$. The determination coefficient is an unbiased measure of the unresolved fluctuations of the estimated quantity.

In addition to this error, in the case where SE-POD is used, it is also possible to calculate a determination coefficient for each POD coefficient separately. This determination coefficient is denoted $R^2_{\mathrm{POD\ single}}$ and defined by:

$$R^2_{\mathrm{POD\ single},i}=1-\frac{\langle(\tilde{a}_i(t)-a_i(t))^2\rangle}{\langle a_i(t)^2\rangle} \tag{II.27}$$

In a more general context, we define the determination coefficients of the estimation of a vector field $\boldsymbol{u}$ by [66]:

$$R^2=1-\frac{\int_\Omega\langle\sum_{i=1}^{n_c}(\tilde{u}_i(\boldsymbol{x},t)-u_i(\boldsymbol{x},t))^2\rangle\mathrm{d}\boldsymbol{x}}{\int_\Omega\langle\sum_{i=1}^{n_c}u_i(\boldsymbol{x},t)^2\rangle\,\mathrm{d}\boldsymbol{x}} \tag{II.28}$$

Since the LSE rests on the covariance between the conditional event and the quantity to estimate, we define the average correlation module by:

$$|\mathrm{corr}(u_i,\mathbf{E})|(\boldsymbol{x})_{\mathrm{av}}=\frac{1}{N}\sum_{i=1}^{N}|\mathrm{corr}(u_i(\boldsymbol{x},t),\mathbf{E}(y_i,t))| \tag{II.29}$$

In addition, to spatially observe the quality of the estimation the normalized mean square error will also be used. This error is defined by:

$$NMSE(u_i(\boldsymbol{x}))=\frac{\langle(\tilde{u}_i(\boldsymbol{x},t)-u_i(\boldsymbol{x},t))^2\rangle}{\langle u_i(\boldsymbol{x},t)^2\rangle} \tag{II.30}$$

At last, SE, and more generally the Least Square Regression, may suffer from overfitting. As defined by Abu-Mostafa et al. [72], overfitting is “fitting the data more than is warranted”. In a simple way, a regression suffers overfitting when the fit (defined by an error criterion, R^2 for instance) on the training set is largely lower than on a validation set (composed of data not used to train the regression). When overfitting occurs, regression becomes more representative of the noise contained in the data of the training set than of the true phenomena in the training set. Several situations may lead to overfitting. One situation is when the training set is not representative of the phenomena that one is trying to fit (the training set does not contain enough samples). Increasing

the complexity of the regression model may also lead to overfitting the training data. As more and more parameters are taken into account, the model may fit the noise of the training data set. Another situation is when the training set suffers from multi-collinearities between the conditional events. In the limiting case of perfect linear dependence between two or more conditional events, then the least square problem become singular. Usually the conditional events exhibit some degree of collinearity which may lead to an ill-conditioned matrix E in equation (II.7). In Least Square Regression, the problem of overfitting can be partly overcome using regularization techniques, for example the Tikhonov regularization (see annex A).

Clark et al. [23] proposed to calculate the Variance Inflation Factor (VIF) to evaluate the level of multi-collinearity between the conditional events. VIF is defined for each conditional event as the inverse of the determination coefficient of the estimation, by LSE, of this conditional event *j* by the others:

$$(\mathrm{VIF})_j = 1/{R_j}^2 \tag{II.31}$$

where:

$${R_j}^2 = \frac{\langle\left(\tilde{\mathbf{E}}(y_j,t) - \langle\mathbf{E}(y_j,t)\rangle\right)^2\rangle}{\langle\left(\mathbf{E}(y_j,t) - \langle\mathbf{E}(y_j,t)\rangle\right)^2\rangle}$$

$\tilde{\mathbf{E}}(y_j)$ refers to the LSE of the conditional event $\mathbf{E}(y_j)$, using the other conditional events. When there is no multi-collinearity between the conditional events, VIF is equal to 1 for all the conditional events. If there are multi-collinearities between the conditional events, the VIF will become larger than 1.

2. The choice of test cases

The initial main objective of the HYBEXCIT project is to investigate and develop methods for the real-time prediction of turbulent flows. Early on, from the literature overview, it was decided to first focus on the SE. The main application of the project was to apply the methods on a frigate helipad in order to help secure shipboard helicopter operations. Thus the final test case of the project was chosen to be the flow downstream a 3D double backward facing step similar to the one studied by Herry [2]. This 3D double backward facing step would be studied experimentally in the L2 wind tunnel. In addition, as milestones of the study, and in order to gradually test the SE, two other experimental test cases were selected. These test cases were chosen at the beginning of the study to make sure of the availability of the experimental facilities.

Several considerations were taken into account to select the two configurations. First, the geometries had to be simple enough to be rapidly installed in the boundary layer wind tunnel. Then, they should correspond to different levels of complexity. In addition, one of the configurations should have been already used with SE in the literature to facilitate the comparison. At last both should be detached flow configurations, since the flow around the frigate helipad considered is a detached flow. For this reason, the first configuration chosen is a 2D backward facing step similar to the one studied by Hudy et al. [13]. The Reynolds number of the experiment was however chosen higher than the one of Hudy's study. The second configuration selected is a 3D configuration, the flow downstream a wall mounted cube. A similar flow has been investigated by Mokhasi et al. [24].

For both configurations, it was decided to run low speed PIV campaigns. The goal of these experimental campaigns was to obtain a large enough database. Indeed, an important number of statistically independent PIV snapshots are required to converge the statistical moments used by SE methods. In the two configurations, the conditional event was chosen to be punctual wall pressure

information. In addition to the low speed PIV campaign, it was decided to run high speed PIV measurements in order to get more insight into the flow dynamics. The high speed PIV also allows conducting a frequency analysis of the flow.

In parallel to the experimental test case, the use of numerical data was also investigated. In order to compare the estimation using experimental or numerical data, a simulation of the 2D backward facing step was conducted. This simulation had to match the experimental flow condition. Numerical simulation was also used in order to provide additional test cases. A very basic simulation of the wake behind a blunt trailing edge was conducted and used to validate the SE implementation. At last to provide a different flow configuration, simulations of the OAT15A airfoil in transonic conditions have also been used. Such a flow, presenting a shock, has not been utilized with SE in the literature.

3. Experimental databases

In this part, the experimental databases exploited and analyzed in this work are described. The experiments were conducted by Jean-Claude Monnier and Jerôme Delva of the ONERA Lille from specifications that were given to them.

3.1. Experimental apparatus

3.1.1. Wind tunnels

In the frame of the ONERA HYBEXCIT project, the two available facilities to conduct experimental studies were the boundary layer wind tunnel and the L2 wind-tunnel of the ONERA Lille center.

3.1.1.1. Boundary layer wind tunnel

The boundary layer wind tunnel of the DAAP/ELV unit at the ONERA Lille is an Eiffel type wind tunnel. The wind tunnel has a square test-section that is 300 mm wide and 300 mm high and remains constant in size over 2.64 m (see Fig. II.1). Its collector has a contraction ratio of 20:1 and is followed by a honeycomb. The wind tunnel is equipped with a variable frequency driven motor capable of producing free-stream velocities of 10 to 40 $m.s^{-1}$ for a turbulent intensity lower than 2%. The velocity can be changed in real time. The total power installed reaches 10 kW. The wind tunnel possesses several optical accesses. It is also equipped with a particle generator (of Multi-Laskins Nozzle type) and a seeding grid (see Fig. II.2). Thus, the wind tunnel is particularly well suited for optical measurements, such as Particle Image Velocimetry and Laser Doppler Velocimetry.

Fig. II.1: Boundary layer wind tunnel

Fig. II.2: Collector and seeding grid

3.1.1.2. L2 wind tunnel

The L2 wind tunnel is initially an open loop wind tunnel. However, it stands within an industrial hall, thus the flow ejected by the diffuser revolves and is re-injected into the collector (see Fig. II.3 and Fig. II.4). The wind-tunnel has a rectangular test-section that is 6 m wide and 2.5 m high and remains constant in size over 13 m in length. The wind tunnel collector has a contraction ratio of 1.83:1 and is followed by a honeycomb. Downstream of the test section, a small diffuser increases the test section height to 2.8 m to leave enough space to fit in 18 fans. The fans are put aside in 3 horizontal lines of 6 fans. The total power installed reached 125 kW. They can be controlled individually and/or in pre-defined sequence with other fans. At full power, the maximum velocity in the empty test section is 13.85 $m.s^{-1}$.

Fig. II.3: The L2 wind tunnel in its hall [2]

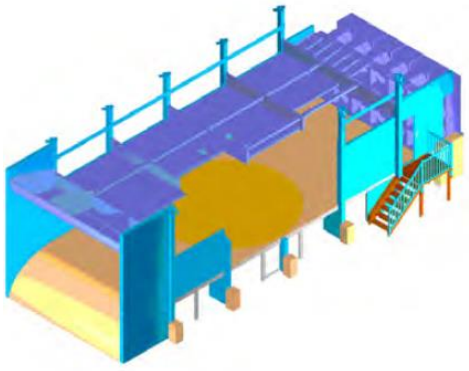


Fig. II.4: Inside view of the L2 wind tunnel [2]

3.1.2. Particle Image Velocimetry

Particle Image Velocimetry is a measurement method that allows accessing the velocity of a flow in a plane or a volume. To record a PIV snapshot, the flow has first to be seeded with particles small enough that they can be considered to follow the flow motion. Then, the area of interest is illuminated twice with a short time interval between each illumination by a double pulse laser and the light diffused by the particles are recorded by cameras. After that the recorded images are split into small interrogation windows and the average displacement of the particles in the window, between the two images, are deduced by post-processing the images through statistical method. Using these information, one gets a snapshot of the velocity from the two recorded images.

Several PIV methods have been developed. The original method is referred to as "2D-2C" and stands for two-dimensional (only a plane is recorded) and two components (only the two components of the velocity in the plane are measured). To access the three components of the velocity it is possible to use "2D-3C" PIV which was introduced by Prasad et al. [73] in 1993. In this method, only a plane is illuminated and recorded but it is viewed from two different directions with separate cameras. It is also possible to access to the velocity in a volume using tomographic PIV (see Elsinga et al. [74])

3.1.2.1. PIV apparatus

For the experimental campaigns, two PIV apparatus were available. The first one is a classical low-speed PIV system. The second one is a high-speed system. For both systems, the "raw" images are processed by a software developed at the ONERA called DAAPPIV [75]. The evaluation of these images is performed with a cross-correlation scheme using Fast-Fourier-Transform (FFT) with multi-pass. The interrogation window size is 32×32 pixels² and a 50% overlap is used.

3.1.2.1.1. Low speed PIV

The low speed PIV uses a twin pulsed 120 mJ Nd:YAG lasers. The highest emission frequency of these lasers is 10 Hz. The recording of the PIV snapshots is carried out by two CCD cameras with 2048×2048 pixels², when 2 components PIV is performed. At full resolution, these cameras can record 5 PIV snapshots per second, limiting the operating frequency of the PIV system to 5 Hz. For 3 components PIV, the recording is carried out by two sCMOS cameras with 2560×2160 pixels².

3.1.2.1.2. High speed PIV

The high speed PIV uses a twin pulsed 20 mJ Nd-YLF lasers. The emission frequency of the lasers can be chosen between 100 Hz and 10 kHz. The PIV snapshots are recorded by two CMOS cameras with 1024×1024 pixels². At full resolution, both cameras can record up to 1 500 PIV snapshots per second.

3.1.3. Wall pressure measurements apparatus

To investigate the use of non-intrusive sensors with the SE. Wall pressure measurements, synchronized with the PIV snapshots, have been recorded. To obtain the wall pressure data, high-speed pressure transducers from Kulite (Model: XCQ-80-5PSID, 5 Psi) have been used.

3.2. 2D Backward facing step experiments

The 2D Backward Facing Step (BFS) height (h) is 30 mm. It has been designed and manufactured by the unit DRIM/DMS of the ONERA (see Fig. II.5). The step is equipped with 17 holes for the pressure transducers (see Fig. II.6). The step occupies an entire side of the boundary layer wind tunnel and has required a modification of the wind tunnel collector. A bakelite plate is mounted in the collector to ensure the connection between the step and the collector (see Fig. II.7).

Fig. II.5: Backward facing step in the boundary layer wind tunnel

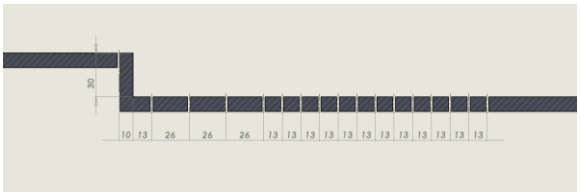


Fig. II.6: Pressure transducers positions

The BFS experiment was conducted for 3 different upstream velocities (U_∞): 20, 25 and 30 m.s^{-1}. The Reynolds numbers (based on the step height) associated with these velocities are approximately $Re_h = 4.05\times10^4$, 5.06×10^4 and 6.08×10^4. For this experiment, only 2 components PIV was carried out. Streamwise and orthogonal to the wall velocities were measured. The velocity in the wind tunnel is measured in real-time by a manometer FCO-150 equipped with a Pitot tube, located downstream of the step. Temperature and absolute pressure, in the wind tunnel, were also measured in real-time to check the stability of the flow condition. The boundary layer has been characterized for each velocity at 10 and 350 mm upstream the step, using a Pitot tube. The results are presented in Table II.1.

Fig. II.7: Wind tunnel collector modifications

Abscissa x(mm)	Upstream velocity U_∞ (m/s)	Boundary layer thickness δ (mm)	Displacement thickness δ^* (mm)	Momentum thickness θ (mm)	Shape factor H
-10	20.02	11.50	1.65	1.25	1.32
-10	24.90	11.50	1.60	1.22	1.31
-10	30.10	11.50	1.65	1.26	1.31

Table II.1: Boundary layer properties

3.2.1. Low speed PIV campaign

The PIV system is schematically represented in Fig. II.10. Picture of the complete set up is displayed in Fig. II.8. Standard optical components were used to form a light sheet approximately 1 mm thick. This light sheet is deviated by a total reflective prism located downstream of the step. This configuration, combined with a common matte black paint, minimizes the reflection from the wall. However, in this configuration, the light sheet impacts orthogonally the step. Therefore, to

make for the intense reflections generated, a special Rhodamine paint was applied to the step. In addition, an interference filter camera lens allowed a limited refection for the camera seeing the step.

The two low speed cameras were employed to form a rectangular acquisition area starting at 1 *h* upstream of the step, extending about 8 *h* downstream of the step and 2 *h* from the wall (see Fig. II.11). The size of each PIV snapshot is 2048×1024 pixel² (half of the available size). The final resolution after post-processing is about 1.1 mm. The PIV acquisition frequency was set to 5 Hz. In addition to the PIV recording, the 17 wall pressure measurements were acquired at a frequency of 20 kHz. This higher frequency is used for the pressure acquisition in order to allow the use of MTD-SE with a large freedom for the choice of delays. It is 200 times the frequency of the main phenomenon (the shedding) and about 70 times the Kelvin-Helmhotz frequency (see Chapter II.4.3). The PIV and the pressure acquisition are synchronized thanks to a common clock, such that 4 000 pressure signals are recorded between every PIV snapshot.

In order to obtain converged statistical moments for the training of the SE, at least 10 000 PIV snapshots were recorded for each velocity. Due to memory constraints on the pressure acquisition, the entirety of the PIV snapshots could not be recorded at once. Thus the acquisition was split into 3 runs for each velocity. For each run, 3 400 PIV snapshots were recorded and the acquisition lasted more than 10 minutes. Between each run, data have to be transferred to the backup device. This transfer took about 10 minutes during which the wind tunnel is kept in operation. The conditions in the wind tunnel were therefore checked in real-time to prevent deviation from the desired conditions.

To sum up, the low speed campaign database comprised data for the 3 upstream velocities (20, 25 and 30 $m.s^{-1}$). For each velocity, 3 separate runs of acquisition have been done. Each run contains 3 400 PIV snapshots and 1.36×10^7 samples for the 17 pressure transducers.

Fig. II.8: General view of the low speed PIV experimental apparatus

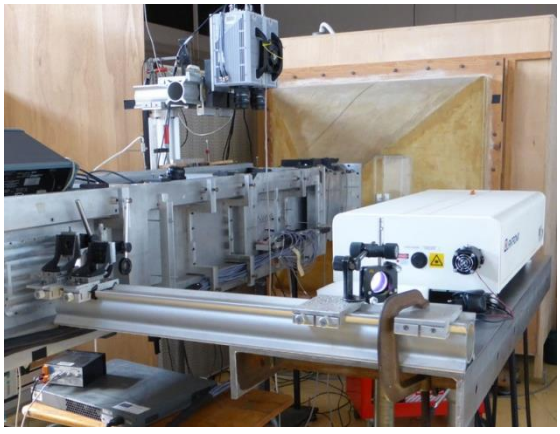


Fig. II.9: General view of the high speed PIV experimental apparatus

3.2.2. High speed PIV campaign

The lighting of the high speed PIV experiment is based on the same setup as in the low speed PIV campaign (see Fig. II.10). In overall, the set-up is unchanged except that the cameras and lasers are replaced by the high speed ones (see Fig. II.9). The size of each PIV snapshot is reduced, compared with the low speed PIV case, and equal to 1024×512 pixels². The resolution of the PIV snapshots, after post-processing, is approximately 1.7 mm. The PIV acquisition frequency was set at 2 kHz (which is higher than the shedding and Kelvin Helmothz frequencies). No synchronized pressure data are available for the high speed PIV campaign.

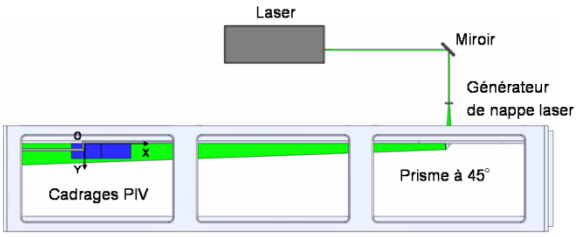


Fig. II.10: Lighting schematics of the PIV for the backward facing step experiment

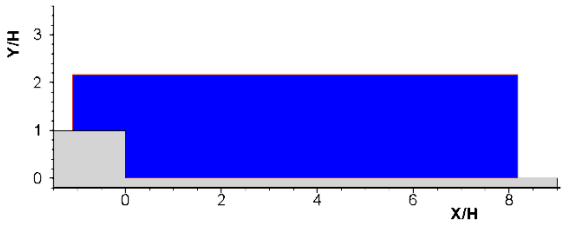


Fig. II.11: PIV recording area for both the low and high speed set up

Owing to the limited memory of the high speed camera, a maximum of 6 144 PIV snapshots could be recorded consecutively. The acquisition of 6 144 PIV snapshots took about 3 s. For each velocity, 4 acquisitions of 6 144 PIV snapshots were performed separately. Between each run, data were transferred to the backup device, which took approximately 20 minutes. During the transfer, the wind tunnel was kept in operation. Therefore, the conditions in the wind tunnel were checked in real-time to prevent deviation from the desired conditions.

To sum up, the high speed campaign database comprised data for the 3 upstream velocities (20, 25 and 30 m.s^{-1}). For each velocity, 4 separate runs of acquisition have been done. Each run contains 6 144 PIV snapshot. No pressure measurement is available.

3.2.3. Experimental results

The average fields obtained for the upstream velocity of 30 m.s^{-1} case using low speed PIV are displayed in Fig. II.12. To estimate the reattachment point, the location of the dividing streamline of the mean flow is used. The reattachment location x_r is about 5.9 h when U_∞ = 30 m.s^{-1}. For U_∞ = 25 m.s^{-1}, the reattachment location is about 5.95 h, and for U_∞ = 20 m.s^{-1}, 5.88 h. These values correspond to the ones found in the literature for comparable Reynolds number [76]. The secondary reattachment point is located around 0.65 h (for the 3 velocities) which is somehow lower than what was found Spazzini et al. (1.1 h at Re_h = 16 000) [77]. Values of the same order of magnitude are obtained from the high speed PIV data for each velocity, and for both reattachment points.

The spectral content of the pressure measurements has also been investigated. Power Spectral Densities (PSD) are computed, in this study, using Welch method and Hamming window (except if stated otherwise). The PSD obtained for the 3 upstream velocity cases and all of the pressure sensors (low speed PIV experiment) are displayed in Fig. II.14. A peak is observed for some sensors (mainly from the 9th sensors (x/h = 4.77)) around a Strouhal number of 0.12. This frequency is associated with the vortex shedding and is consistent with the theoretical value of 0.1 (Dandois et al. [78]). Fig. II.15 shows the frequency wavenumber spectrum obtained for the 30 m.s^{-1} case. A convective velocity is identified at 17 m.s^{-1}, about 0.57 U_∞. This value corresponds to the one expected from the literature (for instance Lee et al. found a convective velocity of 0.6 U_∞ [79]). It corresponds to the passage of large scales vortices.

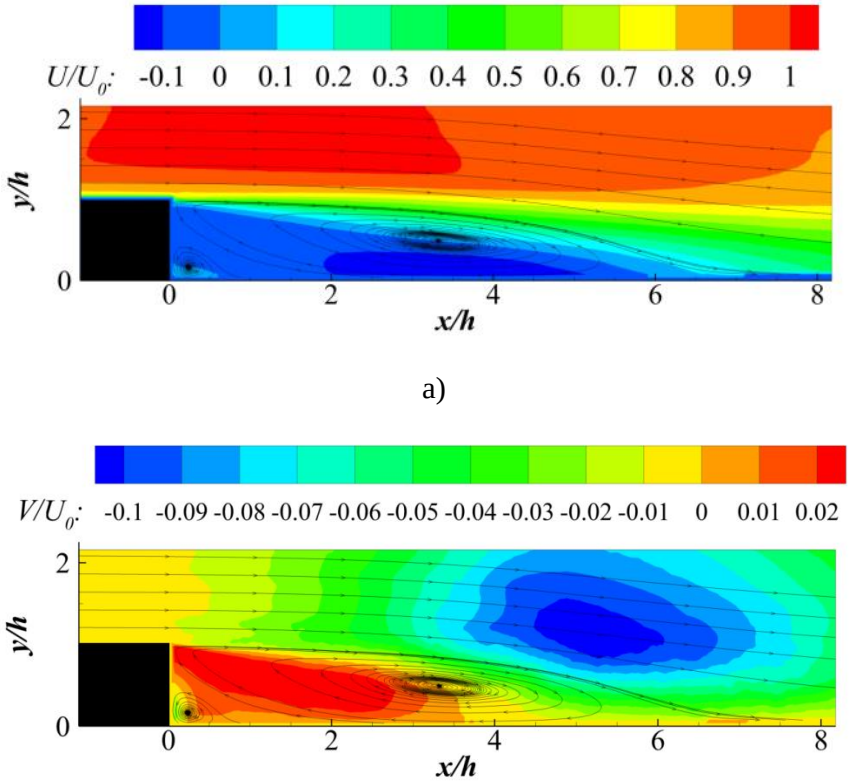


Fig. II.12: Average fields in the 30 m.s^{-1} case from the low speed PIV. a) Streamwise velocity, b) Vertical velocity

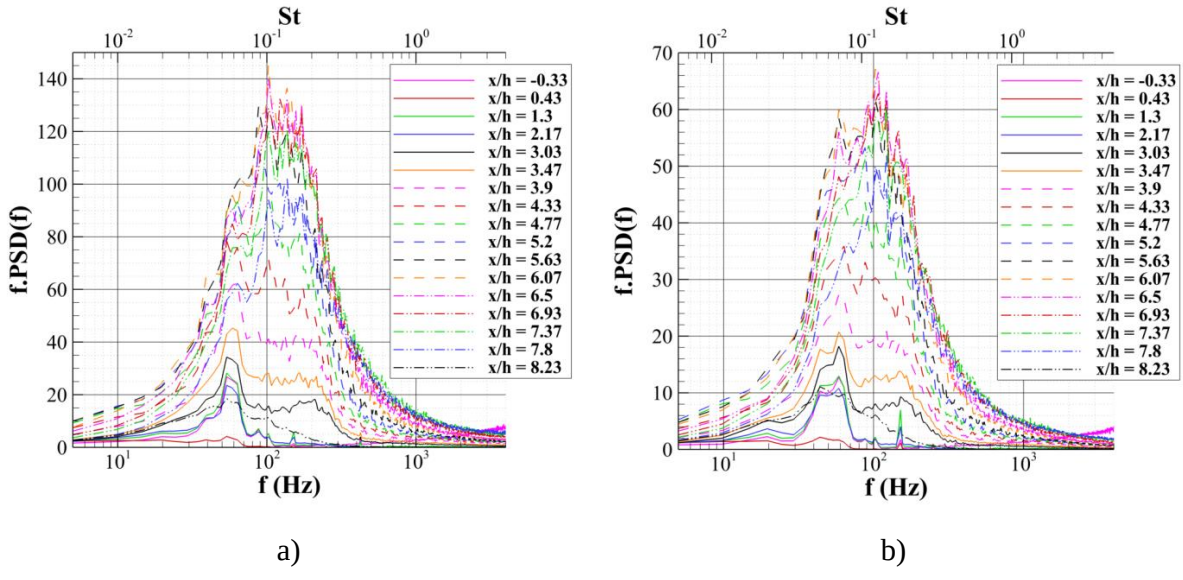


Fig. II.13: PSD of the pressure sensors. a) U_∞ = 30 m.s^{-1}, b) U_∞ = 25 m.s^{-1}

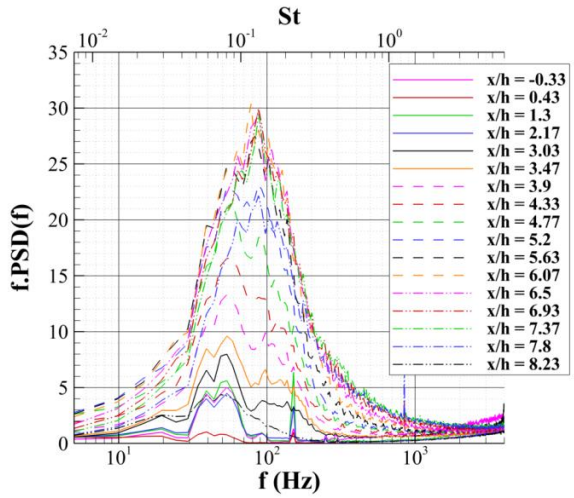


Fig. II.14: PSD of the pressure sensors (U_∞ = 20 m.s^{-1})

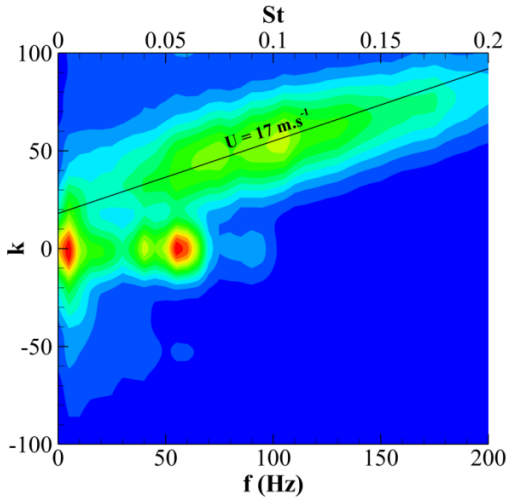


Fig. II.15: Frequency wavenumber spectrum of the 17 pressure sensors (U_∞ = 30 m.s^{-1})

3.2.4. Convergence of the statistics

To evaluate the convergence, and more precisely the uncertainties, of the estimate of the statistical moments used in SE, confidence intervals around the mean are utilized as described by Benedict et al. [70]. For a random variable x, the 95% confidence interval for its mean is defined by:

$$\bar{x} \pm 1.96 \left(\frac{s_x^2}{N} \right)^{0.5} \tag{II.32}$$

where N is the number of statistically independent (uncorrelated) samples of x and s_x is its standard deviation. For a 99% confidence interval, the 1.96 factor should be replaced by 2.58. Since x can be any random variable, it can be the product of 2 or more random variables. Thus to evaluate the uncertainty of the covariance $\langle \mathbf{E}(y_i)\mathbf{E}(y_j) \rangle$ (or any other statistical moment), one can consider $x = \mathbf{E}(y_i)\mathbf{E}(y_j)$ and use the previous formula (II.32).

To determine the number of independent samples, the integral time scales of the pressure sensors have been computed. The integral time scale is defined by:

$$T_i = \int_0^\infty R_P(\tau)\mathrm{d}\tau \tag{II.33}$$

with $R_P(\tau) = \frac{\langle P(t)P(t+\tau)\rangle}{\langle P(t)^2\rangle}$ the autocorrelation function. This function is plotted for the 17 pressure sensors of the U_∞ = 30 m.s^{-1} experiment in Fig. II.16. In practice, the integral is finite. Four strategies are usually found in the literature (see O'Neill et al. [80]): integrate over the entire available domain; if the autocorrelation function has a negative region, integrate only up to the value where the autocorrelation function reaches a minimum; integrate only up to the first zero-crossing; or integrate only up to the value where the autocorrelation function falls under 1/e. Table II.2 summarizes the maximum and minimum values of the integral time scale, using the last strategy (data from the low speed PIV campaign are used, the 2nd pressure sensor is not considered). From the values obtained, for the 3 tested velocities, samples taken at 5 Hz are thus uncorrelated.

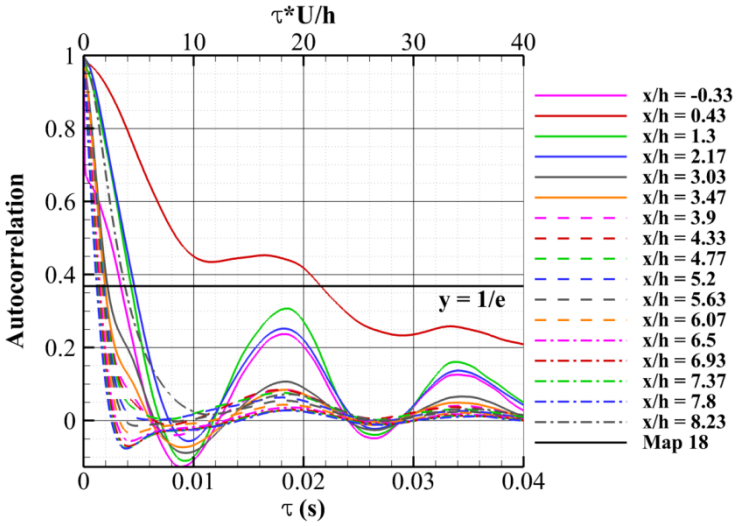


Fig. II.16: Autocorrelation as function of the delay for the 17 pressure sensors (U_∞ = 30 m.s^{-1})

Upstream velocity U_∞ (m.s^{-1})		20	25	30
Minimum T_i	(ms)	1.8	1.5	1.2
	*U_∞/h	1.2	1.25	1.2
Maximum T_i	(ms)	4.3	4.7	4.6
	*U_∞/h	2.9	3.9	4.6

Table II.2: Integral time scales of the pressure sensors (BFS experiments)

3.2.4.1. Low speed PIV experiments

Fig. II.17 shows the values of the covariance $\langle P_i P_j\rangle$ ($i \leq j$) for the 153 permutations (without order) that are used in the SE (U_∞ = 30 m.s^{-1} case). These values are obtained using the 6 800 samples (sampled at 5 Hz) from only 2 of the 3 runs of the U_∞ = 30 m.s^{-1} experiments. Indeed, in order to test the SE, it is necessary to have a training data set and a validation data set. Thus it was decided to use 2 of the 3 runs as training set and 1 as validation set. In the figure, the error bars corresponding to a confidence interval of 95% are also plotted. It appears that the covariance of the pressure sensors is well converged using only data from 2 runs. The relative uncertainties are under 3% (based on 95% confidence interval) for most of the values. Higher relative uncertainties are however obtained when covariance approaches 0. The same conclusion is reached for both U_∞ = 20 m.s^{-1} and

U_∞ = 25 m.s^{-1} cases. The evolution of some values of $\langle P_i P_j \rangle$ as function of the acquisition duration is also displayed in Fig. II.18. If somehow some values seem to be diverging, the global convergence is satisfying. Indeed, the uncertainties for the $\langle P_{10} P_{11} \rangle$ value is only 3% when using the first 6 800 samples (1 360 s) and falls under 2.4% with 10 200 samples.

Some values of $\langle P_i P_j P_k \rangle$ are displayed in Fig. II.19, as well as the 95% confidence interval (in the U_∞ = 30 m.s^{-1} case). This time, all of the 10 200 samples (sampled at 5 Hz) are used. The uncertainties are clearly higher than for the covariance and these statistical quantities appear to be badly converged. The same is also true for the U_∞ = 20 m.s^{-1} and U_∞ = 25 m.s^{-1} cases and even more true for the quantity $\langle P_i P_j P_k P_l \rangle$ in every cases. Thus, when using QSE with these data sets, the QSE may not be representative of the entire phenomena and is likely to suffer from overfitting.

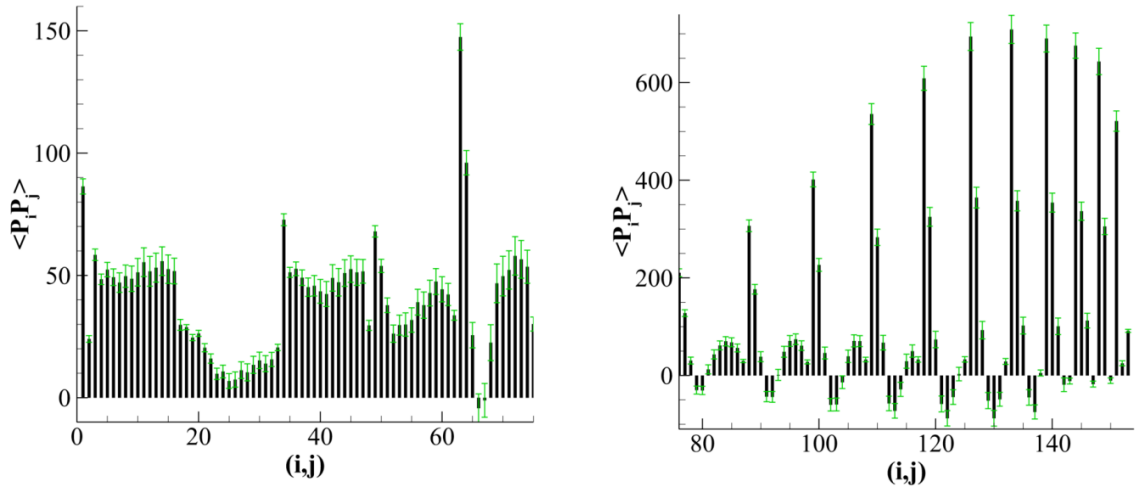


Fig. II.17: Covariance values between the pressure sensors with error bars corresponding to a 95% confidence interval. (i,j) with $i \leq j$.

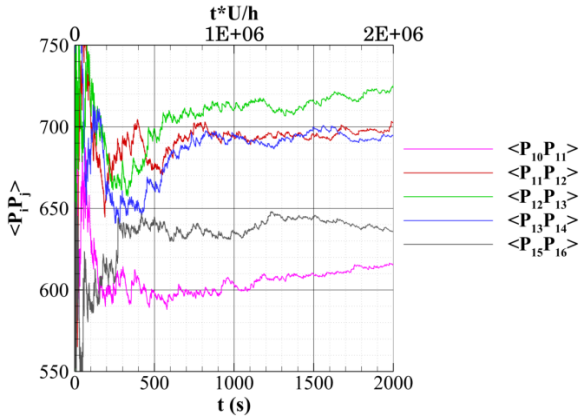


Fig. II.18: Evolution of the estimated covariance (for several combinations of pressure sensors) as function of the acquisition duration (one PIV run lasts for 680 s)

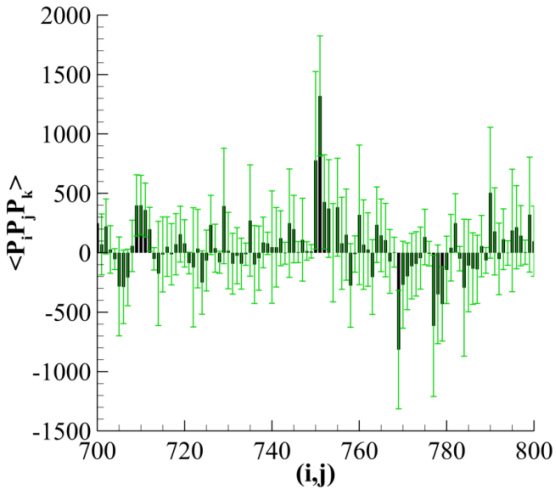


Fig. II.19: Example of $\langle P_i P_j P_k \rangle$ values with their 95% confidence interval (10 200 samples are used)

A similar study (not shown) run for 13 velocity data in the field (see location in Fig. II.20) leads to similar conclusions. The convergence of the covariance $\langle u(x_i) P_j \rangle$ is good, however the uncertainties are higher than for the covariance $\langle P_i P_j \rangle$. The order of magnitude is about 10% of

relative uncertainty based on 95% confidence interval. The statistical moments $\langle u(x_i)P_jP_k\rangle$ do not appear to be sufficiently converged.

Considerations about the convergence of the POD basis are given in annex C.

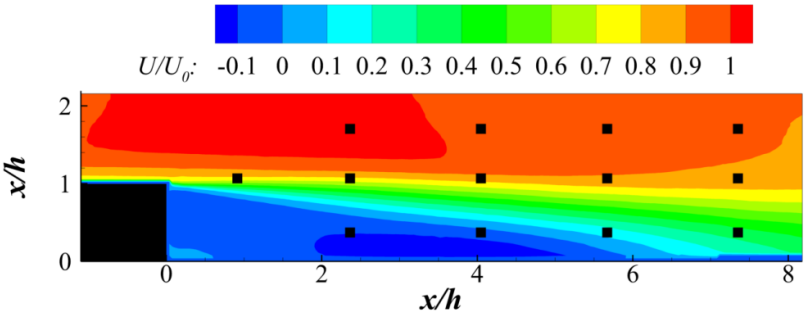


Fig. II.20: Positions of the 13 velocity data, used to study the convergence of the data set (average streamwise velocity field is displayed in background in the 30 m.s^{-1} case)

3.2.4.2. High speed PIV experiments

In these experiments, no pressure data are available. The sampling frequency is 2 kHz, thus the sampling step is shorter than the order of magnitude of the integral time scale (that is about the shedding period). Therefore, the samples cannot be considered statistically independent for the calculation of the uncertainties. In the 30 m.s^{-1} case, the 24 576 available samples represent approximately 1 200 independent samples. This number of independent samples is obtained considering only the largest T_i and thus represents the least favorable hypothesis to estimate the uncertainties. In the two other cases, even less independent samples are available. The covariances $\langle u_iu_j\rangle$ $(i \leq j)$ for the 91 combinations with the 13 streamwise velocity data (see Fig. II.20) are displayed in Fig. II.21 for the 30 m.s^{-1} case. The error bars correspond to a confidence interval of 95%. The statistics seem globally not converged enough. Convergence is not reached for the third and fourth order moments either. The same is observed in the 25 m.s^{-1} and 20 m.s^{-1} cases.

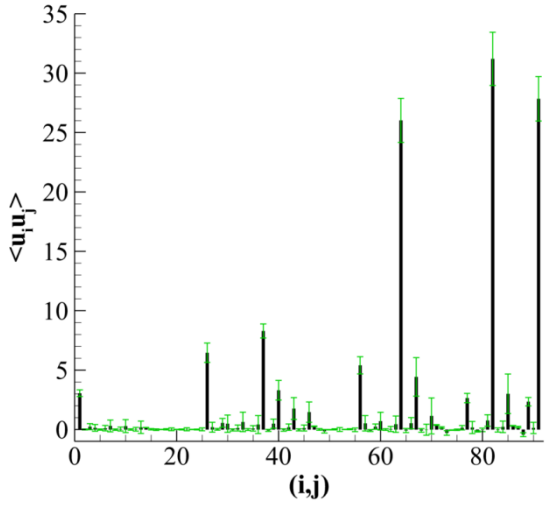


Fig. II.21: Covariance values between the 13 streamwise velocity data with error bars corresponding to a 95% confidence interval. (i,j) with $i \leq j$.

3.3. 3D wall mounted cube experiments

The wall mounted cube has a 30 mm height. The cube is made of glass and fixed to a flat plate. The plate downstream the cube is equipped with 23 slots for the pressure transducers (see Fig. II.22). Indices given to the sensors are described in Fig. II.23 and their positions are detailed in Table II.3. The hole positions were chosen from a pressure sensor location optimization described in Chapter V. Euler simulation of the flow around the cube is used to perform the optimization. The goal is only to get a hint to where the pressure sensors should better be placed than to get a true optimization of their locations. Indeed, the latter would have required running the optimization algorithm on numerical data realistically fitting the experiment.

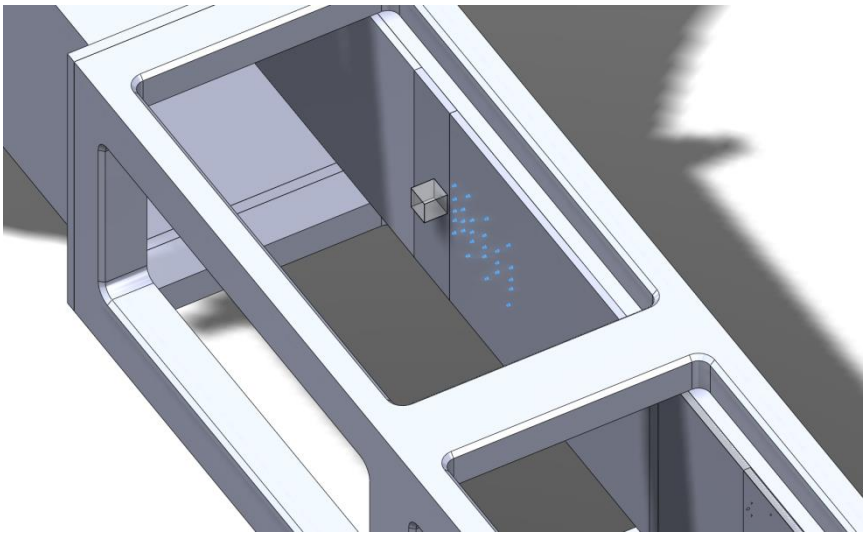


Fig. II.22: Schematics of the wall mounted cube in the boundary layer wind tunnel

As for the BFS experiments, high speed and low speed PIV measurements were conducted. For the low speed case, 6 upstream velocities were tested: 10, 15, 20, 25 and 30 $m.s^{-1}$. The Reynolds numbers (based on the cube height) associated with these velocities are approximately $Re_h = 2.02\times10^4$, 3.04×10^4, 4.05×10^4, 5.06×10^4 and 6.08×10^4. For the high speed case, only 3 velocities were used: 10, 20 and 30 $m.s^{-1}$. For all these experiments 3 components PIV was carried out. During the experiments, the velocity in the wind tunnel is measured in real-time by a manometer FCO-150 equipped with a Pitot tube.

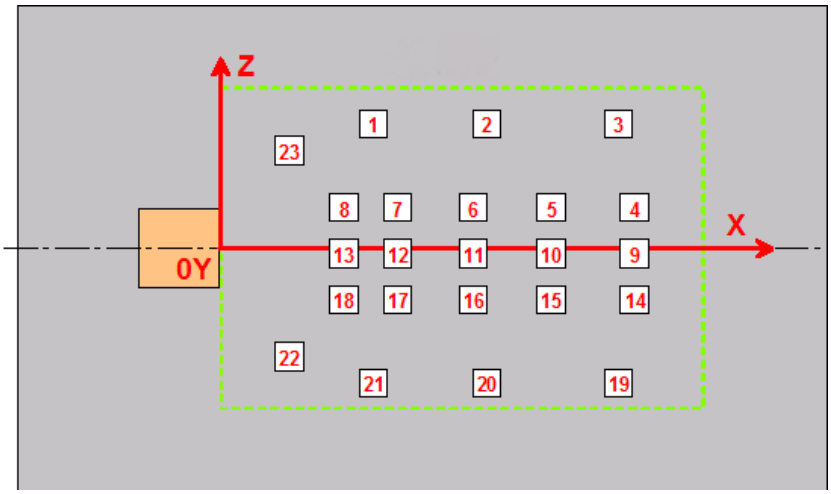


Fig. II.23: Pressure sensor positions

Fig. II.24: Lighting apparatus for the low speed PIV

Fig. II.25: Lighting apparatus for the high speed PIV

For both low speed and high speed PIV campaigns, the lighting strategy was to placed lasers upstream the seeding grid. They were placed far enough of the collector so that they did not disturb the flow in the wind tunnel. Standard optical components were used to generate a light sheet, which entered the wind tunnel from the collector side (see Fig. II.24 and Fig. II.25). Among those components, a converging cylindrical lens was used to change the orientation of the light sheet by rotating it. Two different PIV planes were recorded. One plane is orthogonal to the wall, parallel to the upstream flow and goes through half the cube width (see Fig. II.26). This plane is referred to as the vertical plane. The second plane is parallel to the wall and to the upstream flow, and goes through half the cube height (see Fig. II.27). This plane is referred to as the horizontal plane. Due to the different cameras used in the low and high speed PIV, the size of the horizontal plane recorded is different in each campaign. Field sizes are summarized in Table II.4.

Sensor number	*x/h*	*y/h*	*z/h*
1	1.9	0	1.53
2	3.3	0	1.53
3	4.93	0	1.53
4	5.13	0	0.5
5	4.1	0	0.5
6	3.13	0	0.5
7	2.2	0	0.5
8	1.53	0	0.5
9	5.13	0	-0.07
10	4.1	0	-0.07
11	3.13	0	-0.07
12	2.2	0	-0.07
13	1.53	0	-0.07
14	5.13	0	-0.63
15	4.1	0	-0.63
16	3.13	0	-0.63
17	2.2	0	-0.63
18	1.53	0	-0.63
19	4.93	0	-1.67
20	3.3	0	-1.67
21	1.9	0	-1.67
22	0.87	0	-1.33
23	0.87	0	1.2

Table II.3: Coordinates of the pressure sensors

PIV campaign	Vertical plane	Horizontal plane
Low speed	7.2 h × 2.4 h - h < x/h < 6.2 h 0 < y/h < 2.4 h	6.8 h × 4.4 h 0 < x/h < 6.8 h - 2.2 h < z/h < 2.2 h
High speed	6 h × 2.8 h - h < x/h < 5 h 0 < y/h < 2.8 h	4.2 h × 5.4 h 0 < x/h < 4.2 h - 2.6 h < z/h < 2.6 h

Table II.4: PIV plane sizes

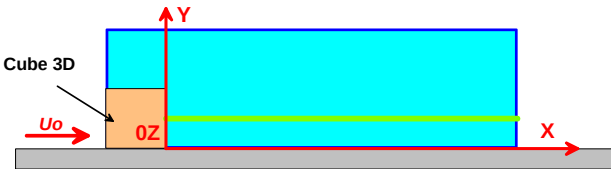


Fig. II.26: PIV vertical plane

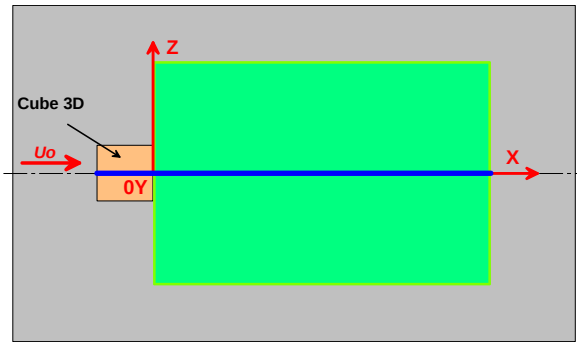


Fig. II.27: PIV horizontal plane

3.3.1. Low speed PIV campaign

The two low speed sCMOS camera are employed and located up and down of the wind tunnel (see Fig. II.28 and Fig. II.29). The final resolution of the PIV snapshots, after post-processing, for both the vertical and horizontal planes, is 1.4 mm. For the same reason as in the BFS experiment, the low speed PIV acquisition was split into runs of 3 400 snapshots recorded at 5 Hz. For each velocity and each plane, 3 runs were performed. Alongside the PIV acquisition, the 23 pressure measurements are recorded at 20 kHz. The PIV and the pressure acquisition are synchronized thanks to a common clock, such that 4 000 pressure signals are recorded between every PIV snapshot.

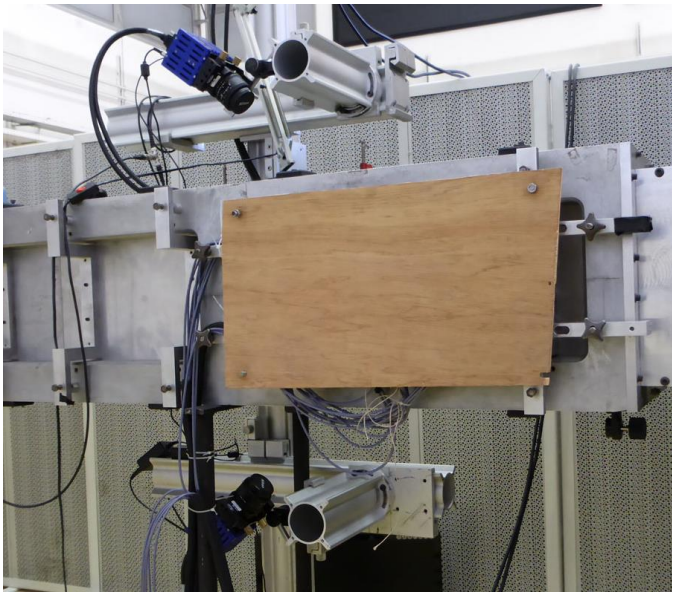


Fig. II.28: Position of the cameras for the low speed PIV vertical plane acquisition

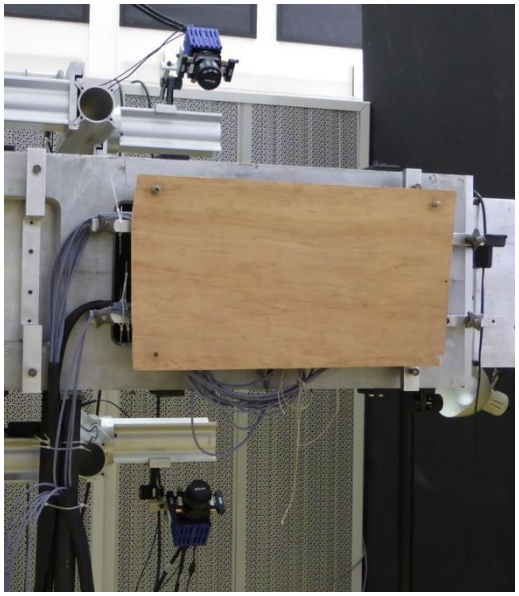


Fig. II.29: Position of the cameras for the low speed PIV horizontal plane acquisition

To sum up, the low speed PIV database comprised data for the two planes and the 6 upstream velocities. For each velocity, 3 separate runs of acquisition have been performed. Each run contains 3 400 PIV snapshots and 1.36×10^7 samples for the 23 pressure transducers.

3.3.2. High speed PIV campaign

For this campaign the two high speed CMOS cameras are employed and positioned similarly to the low speed montage. The resolution of the PIV snapshots, after post-processing, is approximately 2 mm. The PIV snapshots are recorded at 1 kHz (which is at least 10 time the shedding

frequency). Due to the limited memory capacity of the cameras, a maximum of 3 072 PIV snapshots could be recorded consecutively. In an attempt to obtain enough samples to converge the statistical moments used in SE, more runs were performed in the high speed case. The numbers of PIV snapshots recorded for each velocity are related in Table II.5. Alongside the PIV acquisition, the 23 pressure measurements are recorded at 20 kHz. The PIV and the pressure acquisition are synchronized thanks to a common clock, such that 200 pressure signals are recorded between every PIV snapshot.

Upstream velocity ($m.s^{-1}$)	Vertical plane	Horizontal plane
10	16 × 3 072 PIV snapshots	16 × 3 072 PIV snapshots
20	12 × 3 072 PIV snapshots	12 × 3 072 PIV snapshots
30	8 × 3 072 PIV snapshots	8 × 3 072 PIV snapshots

Table II.5: High speed PIV database for the 3D wall mounted cube experiment

3.3.3. Experimental results

The average velocity fields in the vertical plane, obtained for the upstream velocity of 30 $m.s^{-1}$ using low speed PIV, are displayed in Fig. II.30 and Fig. II.31. A main recirculation bubble is observed downstream the cube. It is actually the footprint of an arch vortex (see Sousa et al. [81]). The reattachment lengths x_r/h in the several situations are given in Table II.6. These values are consistent with those of the literature. Martinuzzi et al. [82] found $x_r/h \approx 1.61$ for $Re_h = 10^5$, Lyn et al. [83] found $x_r/h \approx 1.38$ for $Re_h = 2.1\times10^4$ (in a water channel). Tables of values for the reattachment length can be found in Krajnovic et al. [84] and Iaccarino et al. [85]. A roof vortex is also observed (see Sousa et al. [81]), but the flow does not reattach on the cube roof (whatever the upstream velocity). Average velocity fields for the horizontal plane are displayed in Fig. II.32. The arch vortex can be observed. The horseshoe vortex is however under the horizontal plane and is not captured.

U_∞ ($m.s^{-1}$)	10	15	20	25	30
Re_h	2.02×10^4	3.04×10^4	4.05×10^4	5.06×10^4	6.08×10^4
Low speed PIV	1.54	1.43	1.31	1.37	1.37
High speed PIV	1.65	-	1.38	-	1.37

Table II.6: Reattachment length x_r/h obtained in the several configurations

The spectral content of the pressure measurements has been studied. As previously, PSD are computed using the Welch method with Hamming window. The PSD of the 23 pressure sensors are displayed in Fig. II.33 and Fig. II.34 for the U_∞ = 30 $m.s^{-1}$ and 10 $m.s^{-1}$ cases respectively. For all cases a low frequency peak about $St \approx 0.1$ is observed for sensors that are away from x-axis and the most upstream. This frequency corresponds to the vortex shedding occurring at each lateral side of the cube and can concur with values found in the literature (Meinders et al. [86], Yakhot et al. [87]). For some sensors (mainly further downstream, at the center), a second peak is observed at $St \approx 0.2$. This frequency still corresponds to the vortex shedding, but vortices generated at both lateral sides are detected. In some of the PSD displayed, peaks also arise at 50 Hz (peaks at 150, 250 Hz can also be observed), this frequency does not correspond to any physical phenomena and is related to the frequency of the electrical network due to an insufficient isolation of the sensor cables.

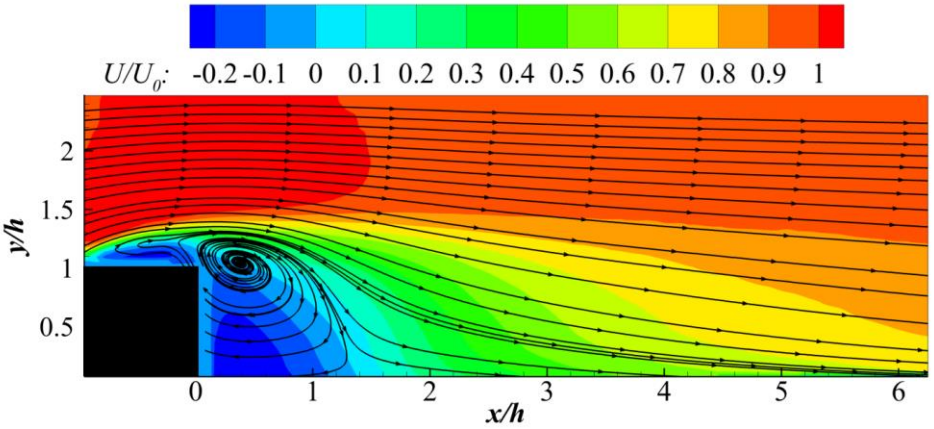


Fig. II.30: Average streamwise velocity fields in the 30 $m.s^{-1}$ case from the low speed PIV (vertical plane)

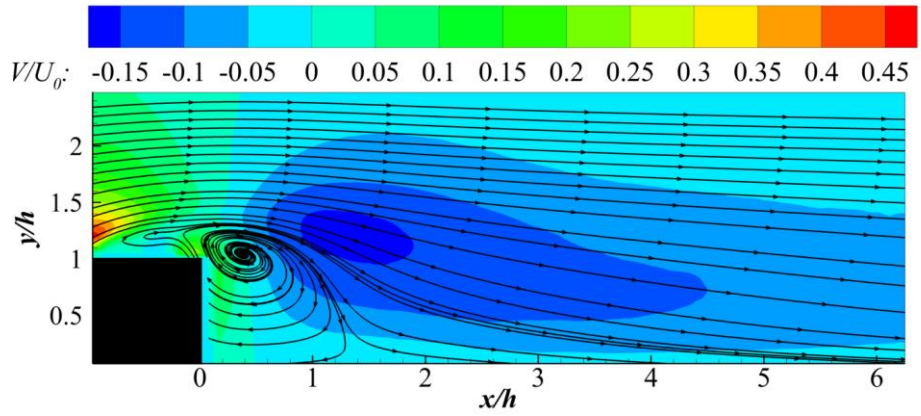


Fig. II.31: Average vertical velocity field in the 30 $m.s^{-1}$ case from the low speed PIV (vertical plane)

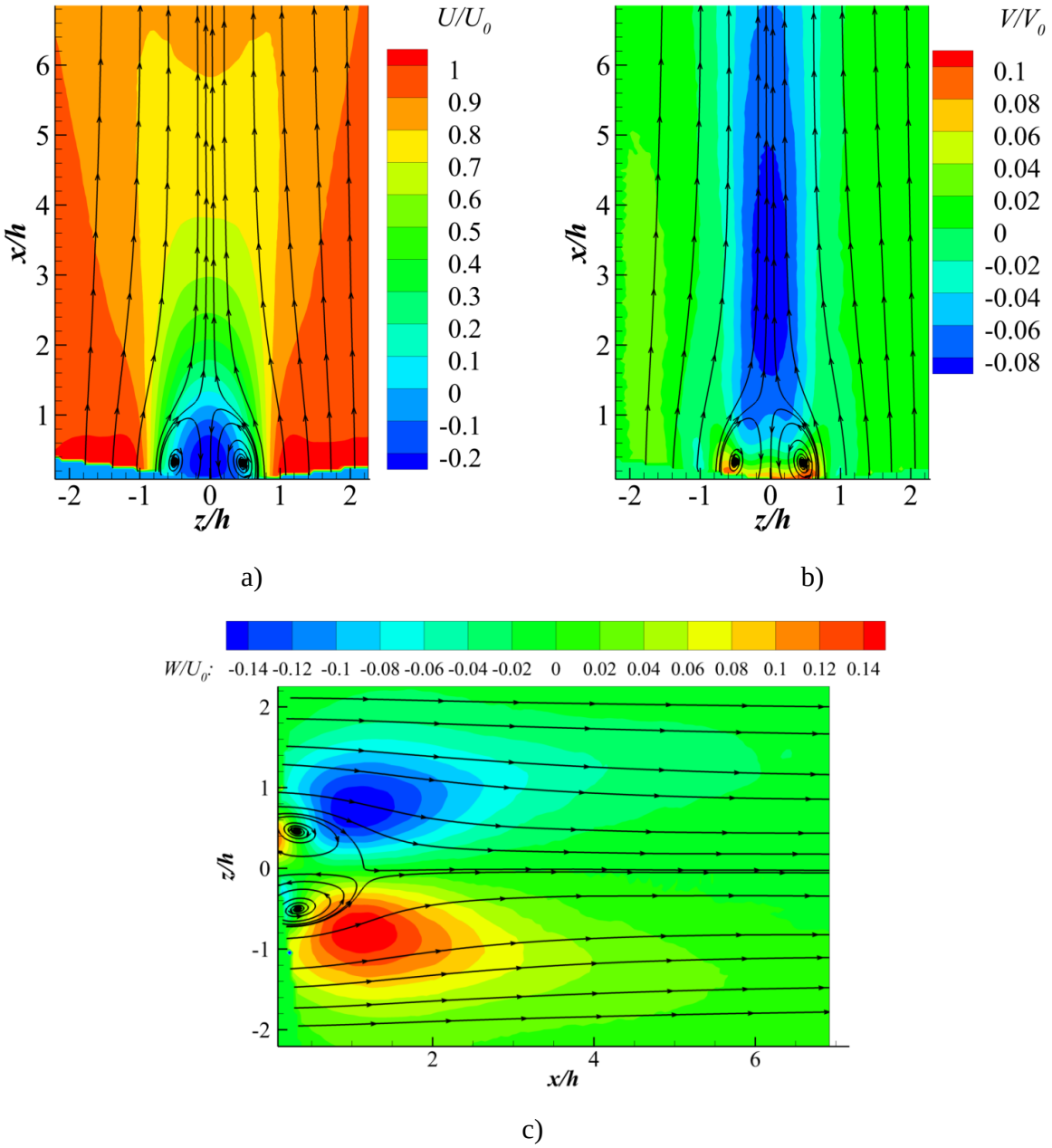


Fig. II.32: Average velocity fields in the 30 m.s^{-1} case from the low speed PIV (horizontal plane). a) Streamwise velocity, b) vertical velocity, c) spanwise velocity

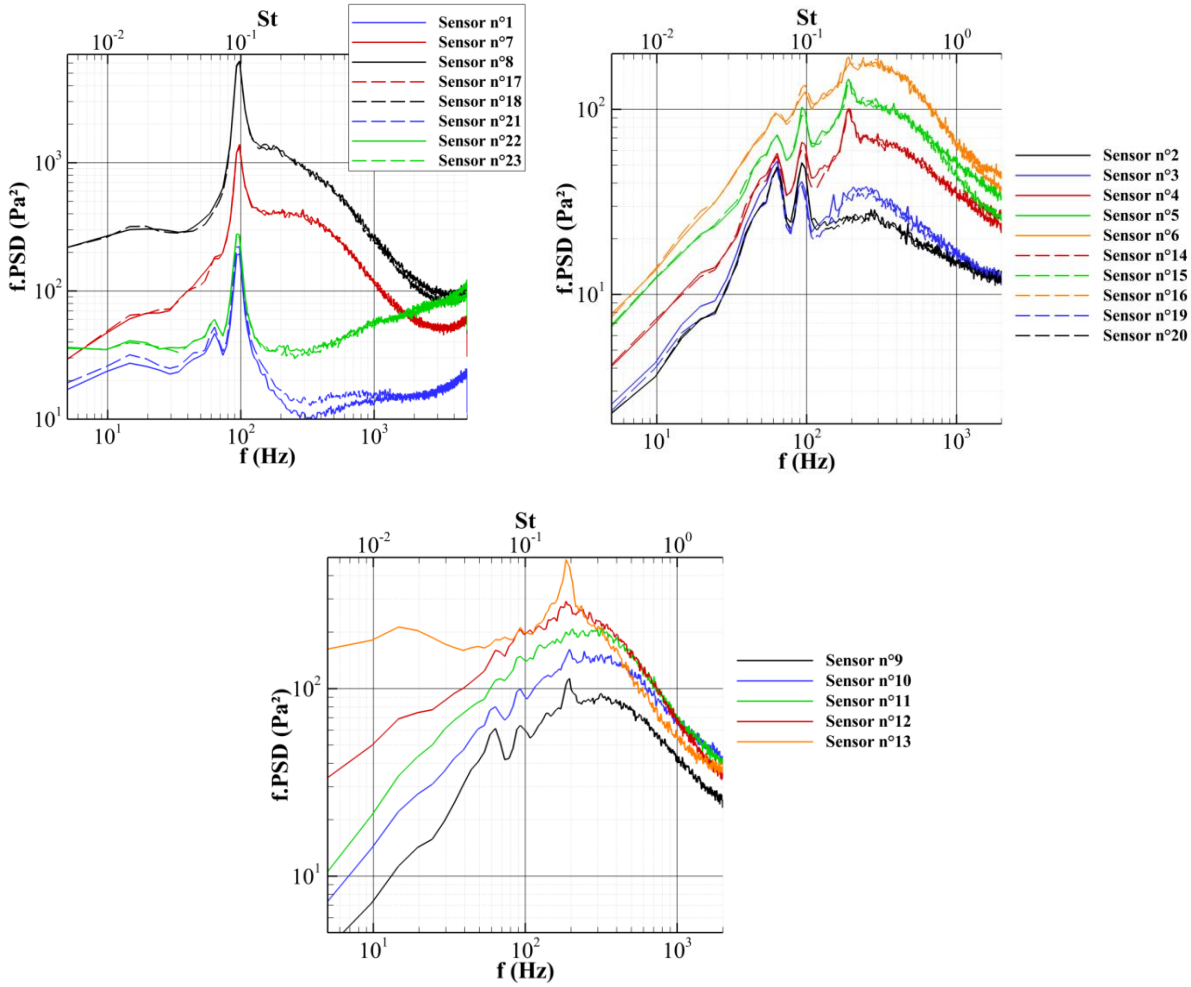


Fig. II.33: PSD of the pressure sensors (U_∞ = 30 m.s^{-1})

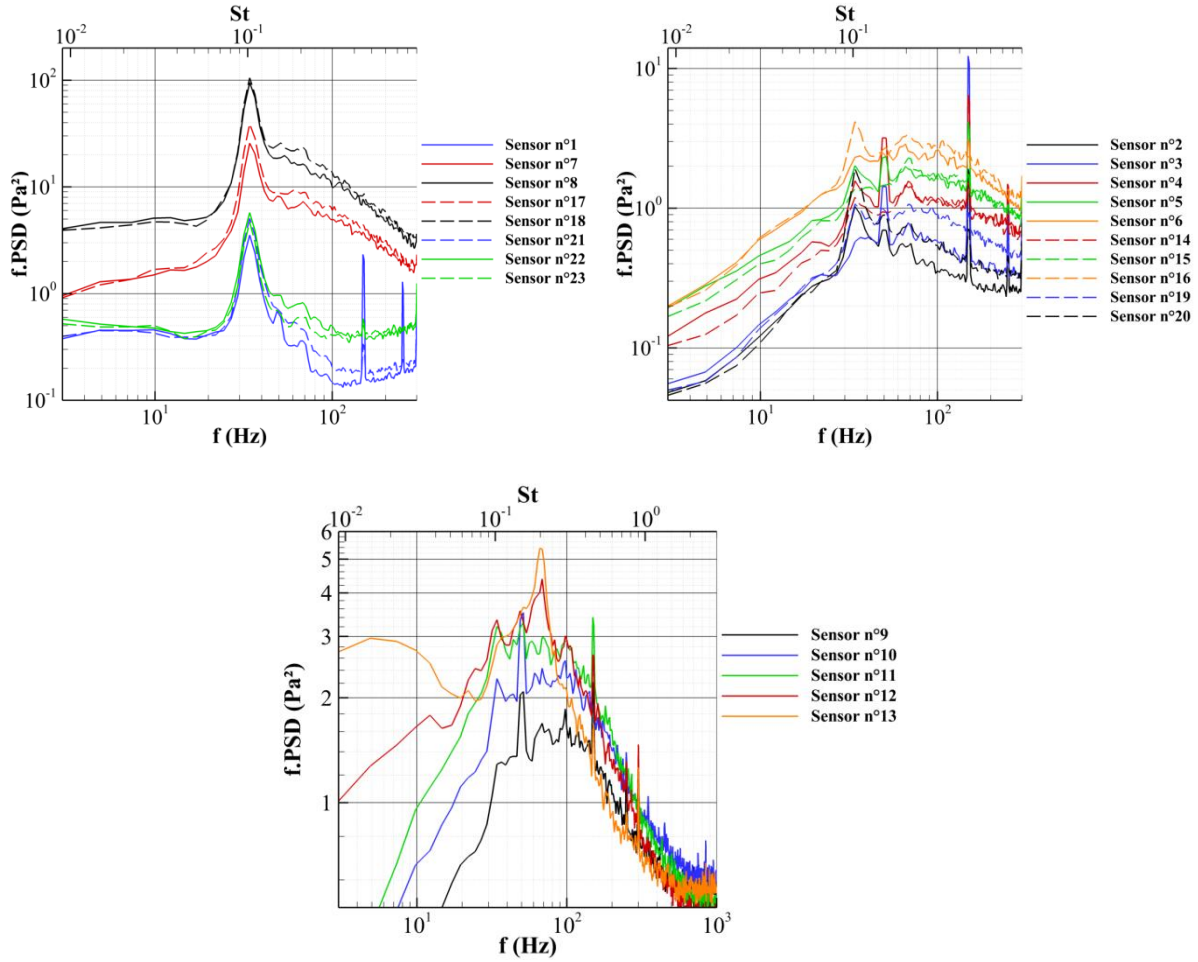


Fig. II.34: PSD of the pressure sensors (U_∞ = 10 m.s^{-1})

3.3.4. Convergence of the statistics

The autocorrelation of the pressure sensors is first investigated and is plotted in Fig. II.35 for the 15 m.s^{-1} case. Using the same methodology as in Chapter II.3.2.4, the extremum values of the integral time scale of the pressure sensors are reported in Table II.7. Even in the 10 m.s^{-1} case, samples taken at 5 Hz are uncorrelated. Once more the order of magnitude of the sampling frequency to obtain statistically independent samples is about the shedding frequency ($St \approx 0.1$).

Upstream velocity U_∞ (m.s^{-1})		10	15	20	25	30
Minimum T_i	(ms)	1.9	3.66	3.2	2.6	2.1
	*U_∞/h	0.63	1.83	2.13	2.17	2.1
Maximum T_i	(ms)	14.3	14.7	10.3	8	8.9
	*U_∞/h	4.77	7.35	6.87	6.67	8.9

Table II.7: Integral time scales of the pressure sensors (cube experiments)

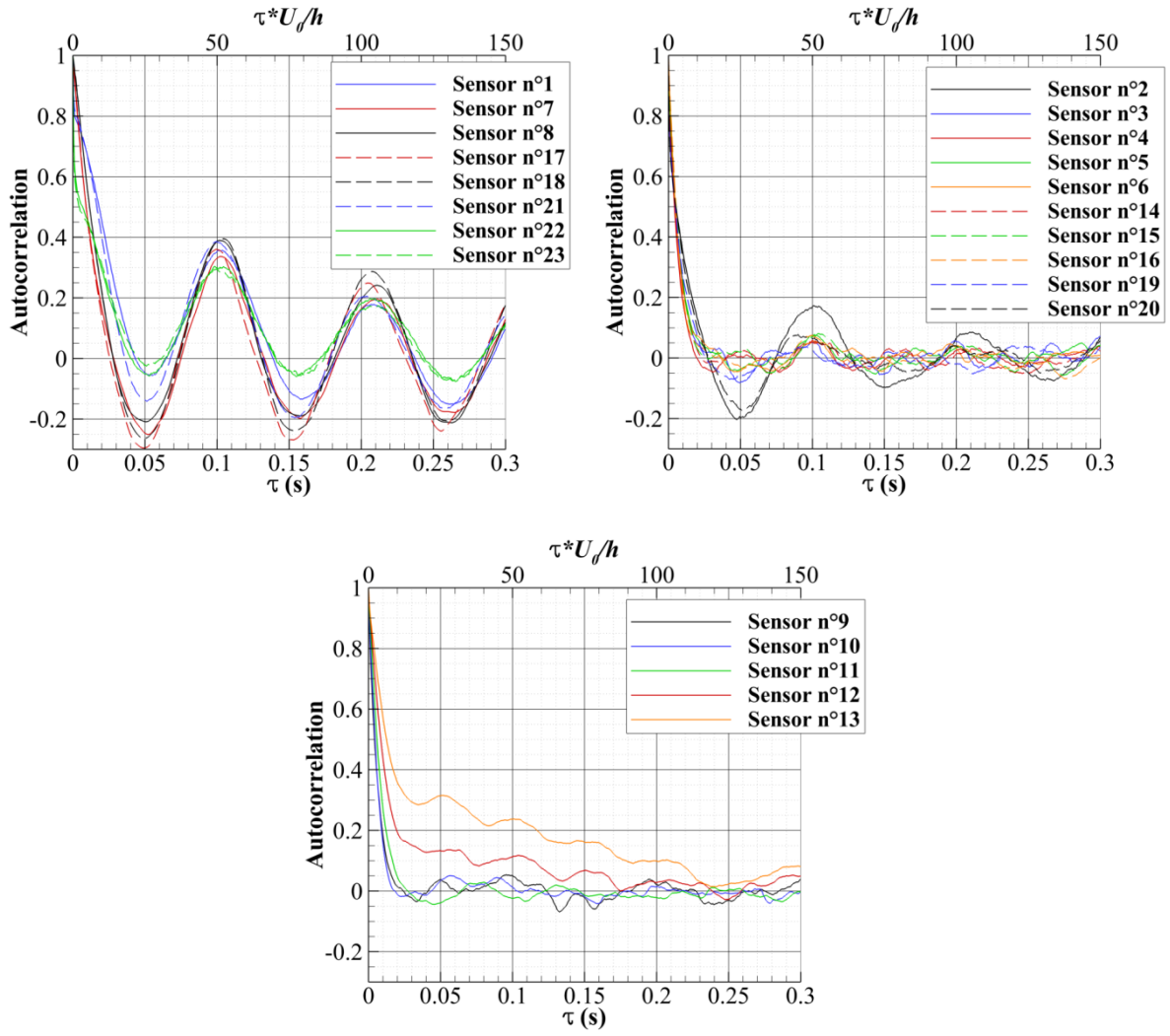
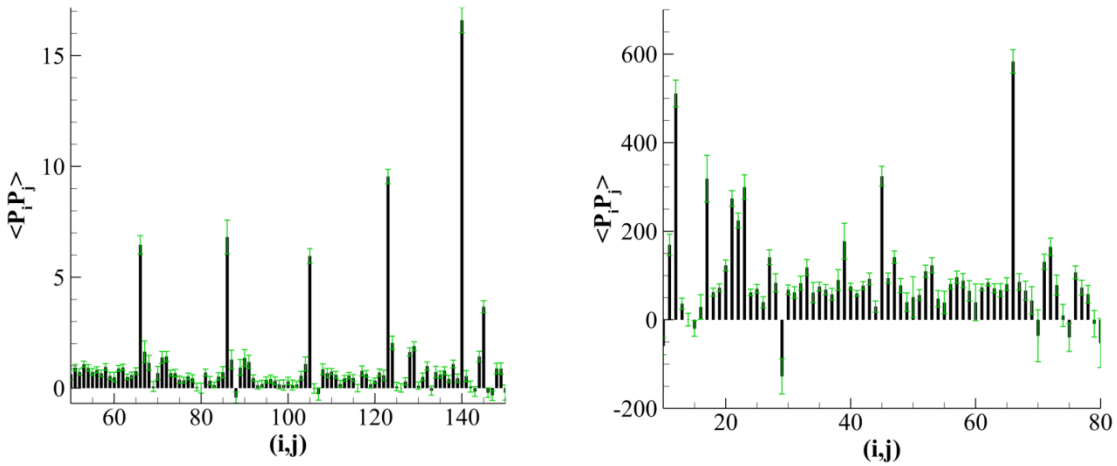


Fig. II.35: Autocorrelation as function of the delay for the 23 pressure sensors (U_∞ = 15 m.s^{-1})

3.3.4.1. Low speed PIV experiments

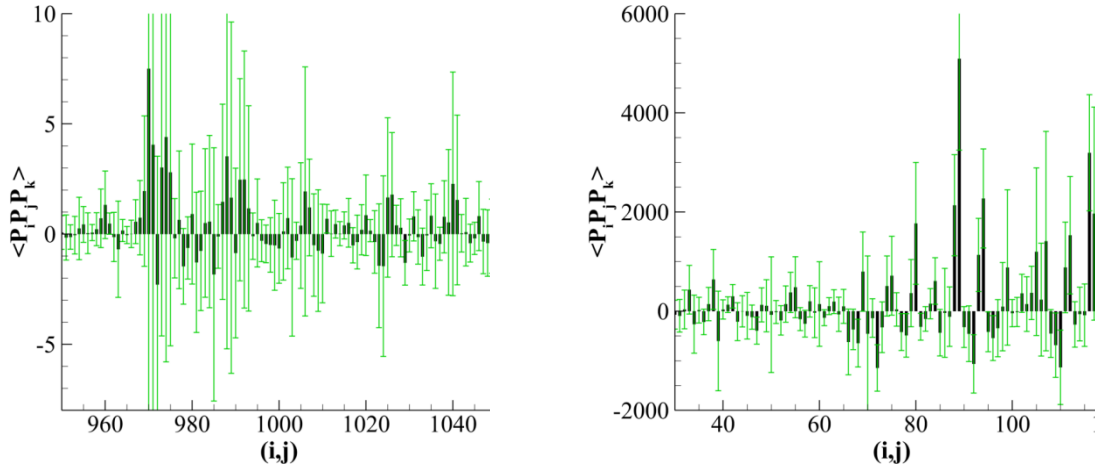
Examples of values of the covariance $\langle P_i P_j \rangle$ ($i \leq j$) are plotted in Fig. II.36 with the error bars corresponding to the 95% confidence interval for U_∞ = 10 and 30 m.s^{-1}. These values are obtained using only 6 800 samples (sampling at 5 Hz) from 2 of the 3 experimental runs available. Most of relative uncertainties are under 10 %, in the U_∞ = 30 m.s^{-1} case, and under 15 %, in the U_∞ = 10 m.s^{-1} case. The convergence is not as good as in the BFS experiments, but we consider it sufficient. Similar conclusions can be drawn for the other upstream velocities.

Fig. II.37 shows some values of the statistical moment $\langle P_i P_j P_k \rangle$. Error bars corresponding to the 95% confidence interval are also plotted (10 and 30 m.s^{-1} cases). From these figures it clearly appears that those statistics are badly converged. For most of the $\langle P_i P_j P_k \rangle$ values, the relative uncertainty is higher than 30% (U_∞ = 30 m.s^{-1}). Similarly, $\langle P_i P_j P_k P_l \rangle$ values do not seem to be sufficiently converged (not shown). And the relative uncertainty is higher than 20% for most of the $\langle P_i P_j P_k P_l \rangle$ values. The same conclusions are reached for all the upstream velocities. Thus the training of the QSE using data from 2 of the 3 runs of the low speed PIV campaign is likely to be unrepresentative of the entire phenomena and to suffer from overfitting.



a) b)

Fig. II.36: Examples of covariance values between the pressure sensors with error bars corresponding to the 95% confidence interval. (i,j) with $i \leq j$. a) U_∞ = 10 m.s^{-1}, b) U_∞ = 30 m.s^{-1}



a) b)

Fig. II.37: Examples of $\langle P_i P_j P_k \rangle$ values with their 95% confidence interval (6 800 samples). a) U_∞ = 10 m.s^{-1}, b) U_∞ = 30 m.s^{-1}

To study the convergence of $\langle u(x_i) P_j \rangle$ and $\langle u(x_i) P_j P_k \rangle$, velocities at 8 positions are extracted for both the vertical and horizontal PIV plane. The extracted data are used to evaluate the convergence of statistical moments between the velocity field and the pressure sensors. The 8 positions are displayed in Fig. II.38. Values of $\langle u(x_i) P_j \rangle$ are shown in Fig. II.39 for the vertical plane, in the 20 m.s^{-1} case and for the streamwise velocity. As previously, only 6 800 samples are used to obtain the figure. While the relative uncertainty is higher than for the covariance between the pressure sensors, the convergence of $\langle u(x_i) P_j \rangle$ is judged qualitatively satisfying. On the contrary, $\langle u(x_i) P_j P_k \rangle$ are found to be badly converged (see Fig. II.40 for instance). These conclusions remain true for the 3 velocity components and the 6 upstream velocity cases.

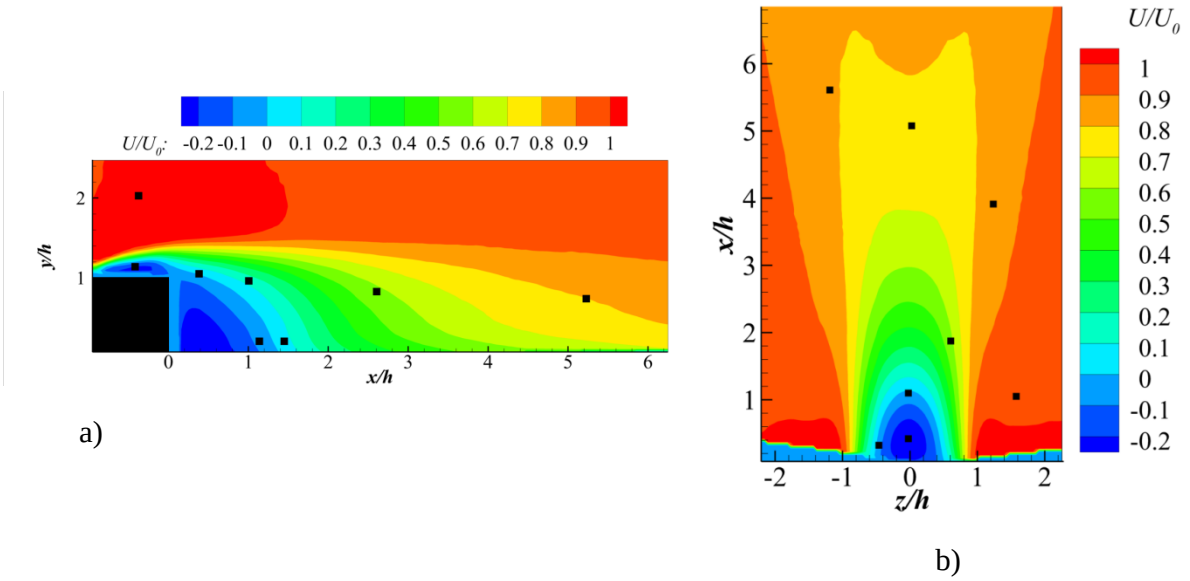


Fig. II.38: Positions of the 8 velocity data points used to study the convergence of statistical moments between the velocity field and pressure sensors. a) Vertical plane, b) horizontal plane. (Average field from the 30 m.s^{-1} is displayed)

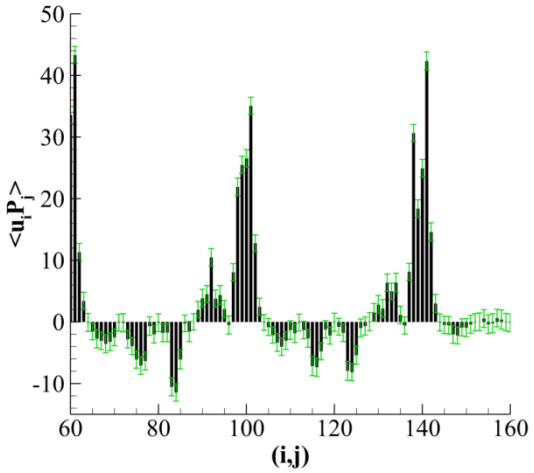


Fig. II.39: Examples of $\langle u(x_i)P_j \rangle$ values with their 95% confidence interval (vertical PIV plane in the 20 m.s^{-1} case)

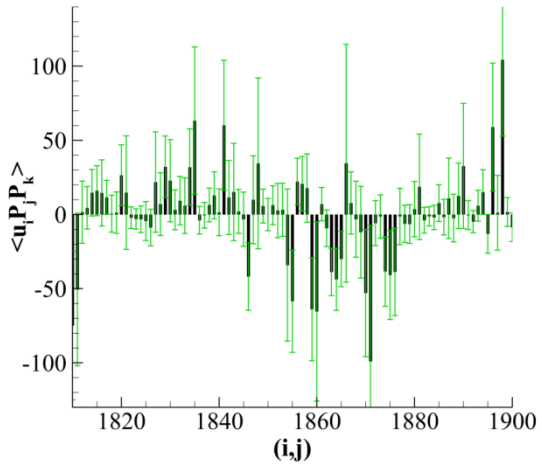


Fig. II.40: Examples of $\langle u(x_i)P_jP_k \rangle$ values with their 95% confidence interval (vertical PIV plane in the 20 m.s^{-1} case)

In addition to studying the statistics uncertainties, as proposed by Benedict et al. [70], it is also possible, for the statistical moments between the pressure sensors only, to compare values obtained from each PIV plane. Indeed, measurements are done separately for each PIV plane. This comparison shows that the convergence of the covariance $\langle P_iP_j \rangle$ is reached. Indeed, an excellent agreement is obtained in every cases (see for instance Fig. II.41). For higher order moments, differences between the values become larger, thus confirming their poor convergence (see Fig. II.42).

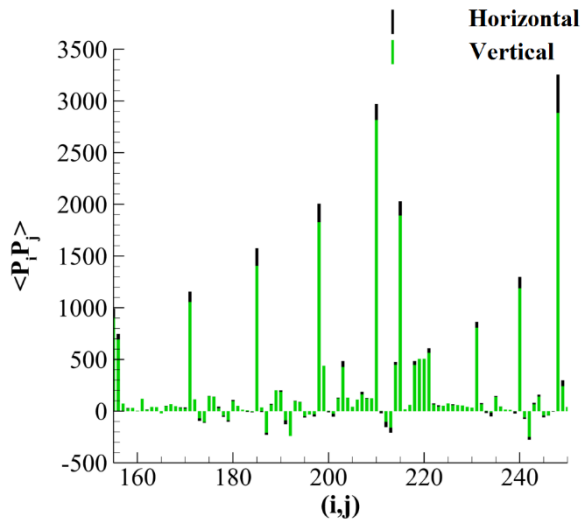


Fig. II.41: Comparison of $\langle P_i P_j \rangle$ values obtained from the training set of the horizontal and vertical PIV plane (30 m.s^{-1} case)

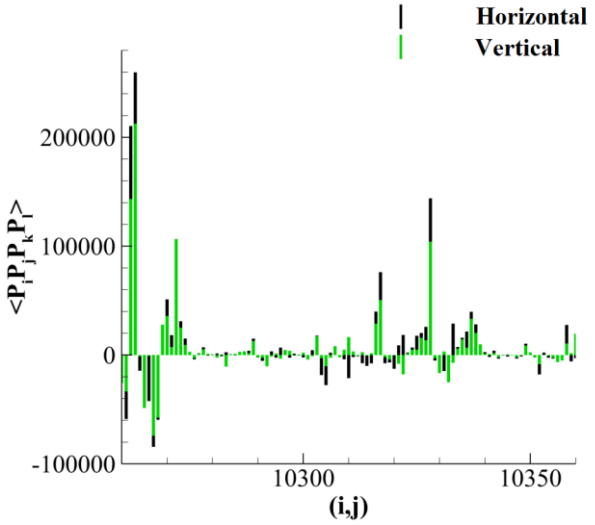


Fig. II.42: Comparison of $\langle P_i P_j P_k P_l \rangle$ values obtained from the training set of the horizontal and vertical PIV plane (30 m.s^{-1} case)

3.3.4.2. High speed PIV experiments

As previously explained, in an attempt to obtain relatively converged statistical moments using the high speed PIV database, more samples were recorded during these experiments (see Table II.5). When the upstream velocity is 30 m.s^{-1} the largest integral time scales is about 0.01 s. Thus, the high speed PIV sampling rate being 1 kHz, about 1/10 samples can be considered statistically independent as the most constraining assumption. For U_∞ = 30 m.s^{-1}, it is decided to use 18 432 samples (from 6 runs) to form a training set, the remaining 6 144 samples forming the validation set. Thus the training set contains approximately 1 843 uncorrelated samples (N in equation (II.32)). Examples of values for the covariance $\langle P_i P_j \rangle$ are displayed in Fig. II.43 with the 95% confidence interval. The covariance between the pressure sensors seems to be qualitatively well converged. However, quantitatively, only about one half of the data has a relative uncertainty under 10 %. $\langle P_i P_j P_k \rangle$ and $\langle P_i P_j P_k P_l \rangle$ are not correctly converged (see Fig. II.44 and Fig. II.45).

To study the convergence of $\langle u(x_i) P_j \rangle$ and $\langle u(x_i) P_j P_k \rangle$, once more, velocities at 8 and 7 locations are extracted for the vertical and horizontal PIV plane respectively (the horizontal plane in the high speed PIV experiment being shorter than in the low speed one, only the 7 locations the closest to the cube are available). Fig. II.46 shows values of $\langle u(x_i) P_j \rangle$ when using the spanwise velocity data of the horizontal PIV plane. The cross-covariance is sufficiently converged, but relative uncertainties remain large with only one half of the values with a relative uncertainty lower than 50 %. This is due to the fact that a large amount of $\langle u(x_i) P_j \rangle$ converged to values close toward 0. Similar observations are obtained with the streamwise and vertical velocity in both PIV plane. The convergence of $\langle u(x_i) P_j P_k \rangle$ is however very poor.

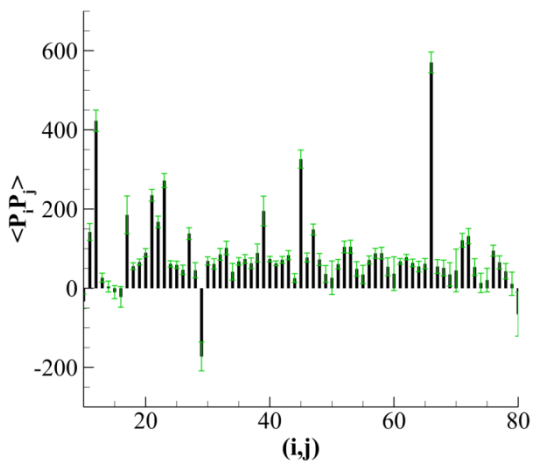


Fig. II.43: Examples of covariance values between the pressure sensors with error bars corresponding to the 95% confidence interval (30 m.s^{-1} case)

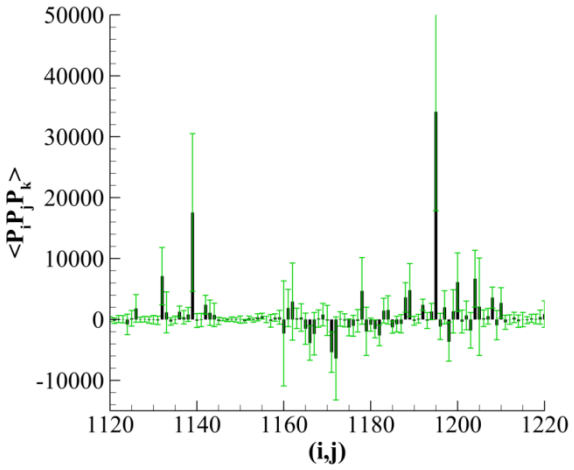


Fig. II.44: Examples of $\langle P_i P_j P_k \rangle$ values between the pressure sensors with error bars corresponding to the 95% confidence interval (30 m.s^{-1} case)

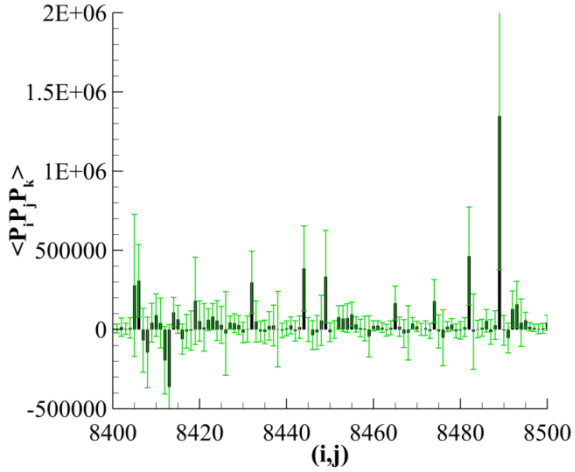


Fig. II.45: Examples of $\langle P_i P_j P_k P_l \rangle$ values between the pressure sensors with error bars corresponding to the 95% confidence interval (30 m.s^{-1} case)

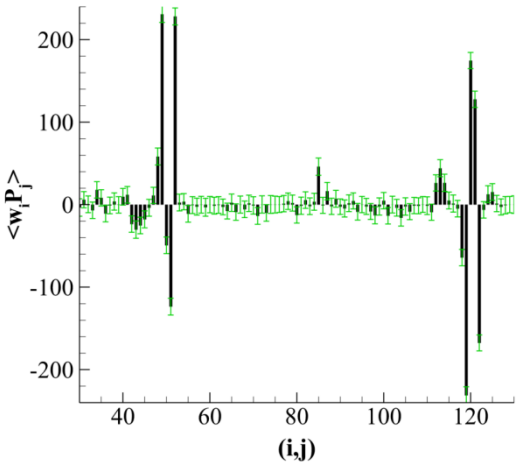


Fig. II.46: Examples of $\langle u(x_i) P_j \rangle$ values with their 95% confidence interval (data from horizontal PIV plane, 30 m.s^{-1} case)

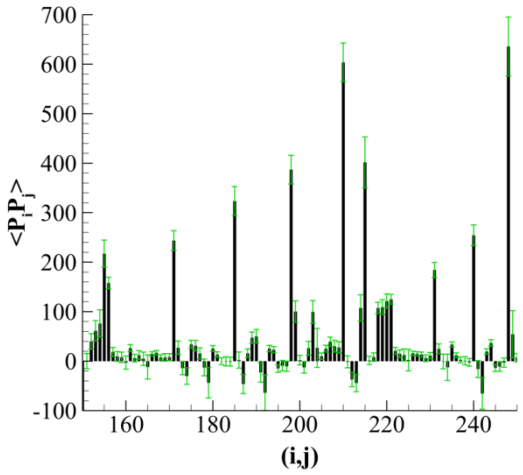


Fig. II.47: Examples of covariance $\langle P_i P_j \rangle$ values with error bars corresponding to the 95% confidence interval (20 m.s^{-1} case)

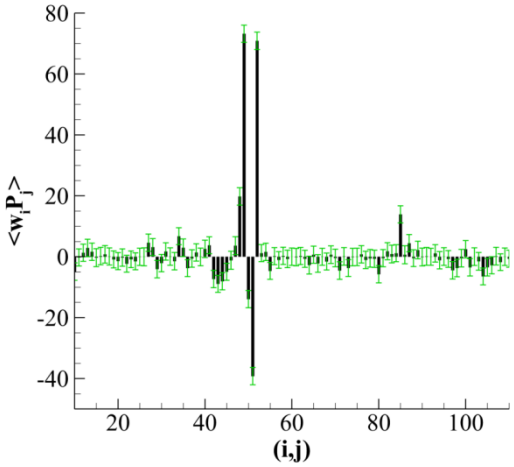


Fig. II.48: Examples of $\langle u_i P_j \rangle$ values with error bars corresponding to the 95% confidence interval (20 m.s^{-1} case)

For the 20 m.s^{-1} upstream velocity, 10 of the 12 runs are used to form the training set. The remaining two give the validation set. In this situation, the training set contains 30 720 samples, thus about 2047 uncorrelated samples. The convergence study of this training set leads to the same conclusions that for the 30 m.s^{-1} case. Fig. II.47 and Fig. II.48 are given as illustrations.

At last, in the 10 m.s^{-1}, 12 of the 16 runs are used to form the training set. The training set therefore contains 36 864 samples, corresponding to approximately 1 230 shedding periods and uncorrelated samples. Same conclusions as previously are reached, which is illustrated in Fig. II.49 and Fig. II.50.

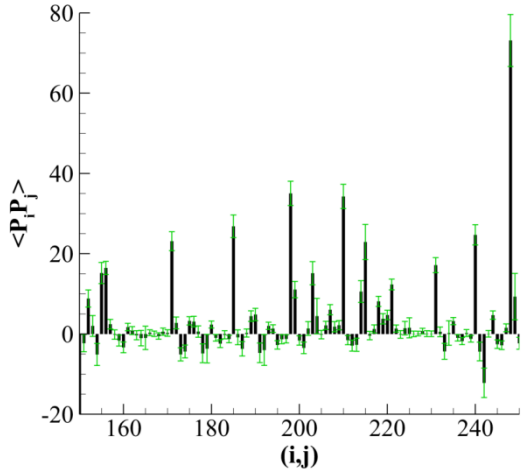


Fig. II.49: Examples of covariance $\langle P_i P_j \rangle$ values with error bars corresponding to the 95% confidence interval (10 m.s^{-1} case)

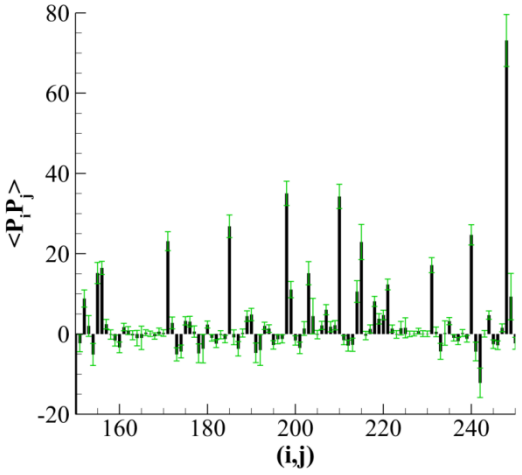


Fig. II.50: Examples of $\langle u_i P_j \rangle$ values with error bars corresponding to the 95% confidence interval (10 m.s^{-1} case)

As previously, it is also possible, for the statistical moments between the pressure sensors only, to compare values obtained for each PIV plane. This comparison confirms the previous conclusions. Convergence of the covariance $\langle P_i P_j \rangle$ is reached and an excellent agreement is obtained in every cases between the values obtained using data from the horizontal PIV plane experiment and

the ones from the vertical PIV plane experiment (see for instance Fig. II.51). For higher order moments, differences between the values become much larger, thus confirming their poor convergence (see Fig. II.52).

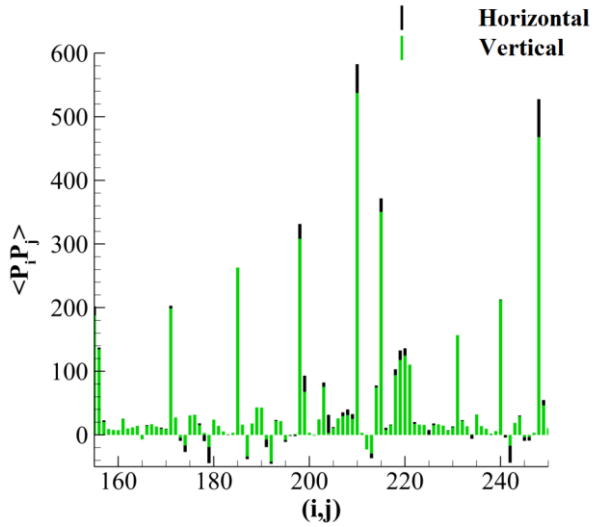


Fig. II.51: Comparison of $\langle P_i P_j \rangle$ values obtained from the training set of the horizontal and vertical PIV plane (30 m.s^{-1} case)

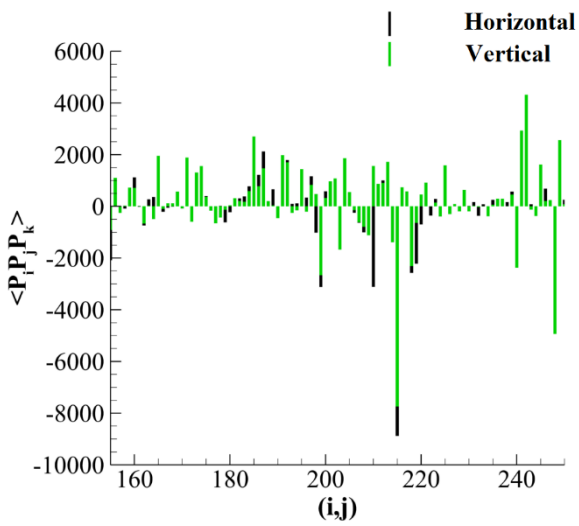


Fig. II.52: Comparison of $\langle P_i P_j P_k \rangle$ values obtained from the training set of the horizontal and vertical PIV plane (30 m.s^{-1} case)

4. Numerical databases

All simulations used in this work have been performed using the elsA software (ensemble de logiciels pour la simulation en Aérodynamique / set of software for aerodynamic simulation) of the Onera. This software is able to perform numerical simulations of external and internal flows. It also covers many aspects of Computational Fluid Dynamics, such as aerodynamics, aeroacoustics, aerothermal and aeroacoustic couplings. With this software it is possible to perform simulations using Euler, RANS, URANS equations and DNS, ZDES or LES methods for instance. The system of equations is solved by a cell centered finite-volume method. For more details about elsA, one can refer to Cambier et al. [88]. Except for the simulations of the flow around the OAT15A airfoil, that comes from the work of Brunet et al. [89] [90], these simulations have all been personally performed.

4.1. Wake behind a blunt trailing edge simulation

In order to validate the implementation of the Stochastic Estimation (and its several extensions), a simple flow field was selected for which the SE was expected to well perform. As seen in Chapter I, SE has already been applied for the reconstruction and prediction of blunt trailing edge wake flows (see Tu et al. [65] and Clark et al. [23] for instance). For such flow, at the Reynolds number they used, most of the TKE was contained in a few POD modes (Tu et al. used 7 POD modes, Clark et al. used 2 POD modes). And SE from wall pressure measurements was able to reproduce with good fidelity the time evolution of those POD modes. Especially, the first two POD modes (which form the shedding) were predicted with high accuracy. R^2 about 85% is reached by Clark et al. using 3 sensors and MTD-LSE-POD.

The main goal being to validate the SE implementation, it was decided to conduct a very simple simulation, without much regard to the physical meaning and consistency with real experimental data. The aim was to quickly obtain a simple data set that could be used rapidly. Thus, a 2D inviscid simulation of a blunt trailing edge was performed. The blunt trailing edge height h is 12.7 mm (which is the same height as in the experiment conducted by Tu et al.). The upstream velocity U_∞ = 10 m.s^{-1}. The simulation was run using the elsA software with a structured mesh of about 31 000 cells. The mesh is displayed in Fig. II.53 and the several blocks can be seen. The simulated domain is about 4 h upstream the trailing edge, 16 h downstream and 12 h vertically. All boundary conditions are non reflective ones, except for the flat plate which is treated as an inviscid wall. The time integration scheme is the Gear one, using 9 sub-iterations and a time step of 2.5 µs. The Jameson scheme was used for the space discretization.

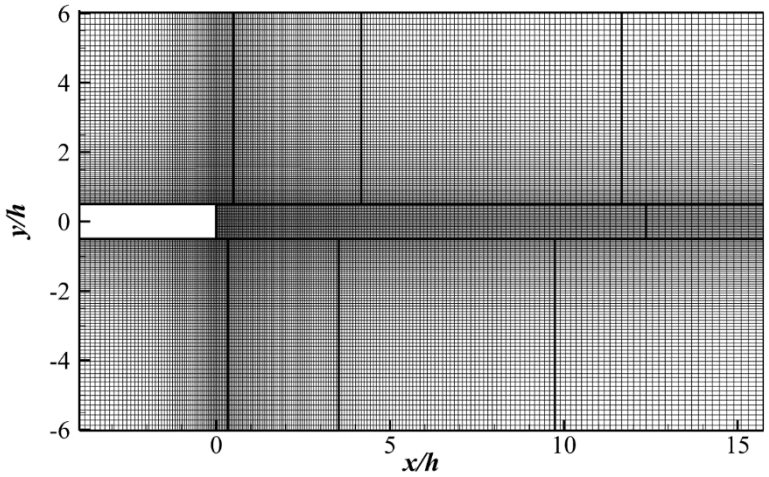


Fig. II.53: Mesh of the blunt trailing edge wake Euler simulation

The shedding frequency observed from the simulation results is about 170 Hz (St ≈ 0.22). This frequency is consistent with the expected Strouhal number of 0.2 and the experimental values of Tu et al. (~0.27) and Clark et al. (~0.23). 1 860 snapshots, sampled at 80 kHz, were extracted, representing around 4 shedding periods. The two velocity components, as well as the static pressure, are extracted in the snapshots. These snapshots form the available numerical database for this case. The average streamwise velocity field is displayed in Fig. II.54. The streamlines show that there are indeed two vortex forming just downstream the trailing edge. An instantaneous streamwise velocity field is shown in Fig. II.55. The von Kármán vortex street is clearly observed.

Euler simulation gives perfectly periodic and reproducible results. It is not a stochastic process. Therefore, the knowledge of one period is enough to obtain the averages required by SE. In such situation, it was considered that the use of about 400 samples was enough to get a correct estimation of all averages. Precisely, a shedding period spans 465 samples and thus 465 snapshots are used as training set. The remaining 1 395 snapshots make the validation set (about 3 shedding periods). Since the sample frequency is not a multiple of the shedding frequency, the 465 samples obtained for each shedding period are slightly different. It remains a poor way to evaluate the SE performances, but the main goal, when using this database, is to validate the implementation more than anything else.

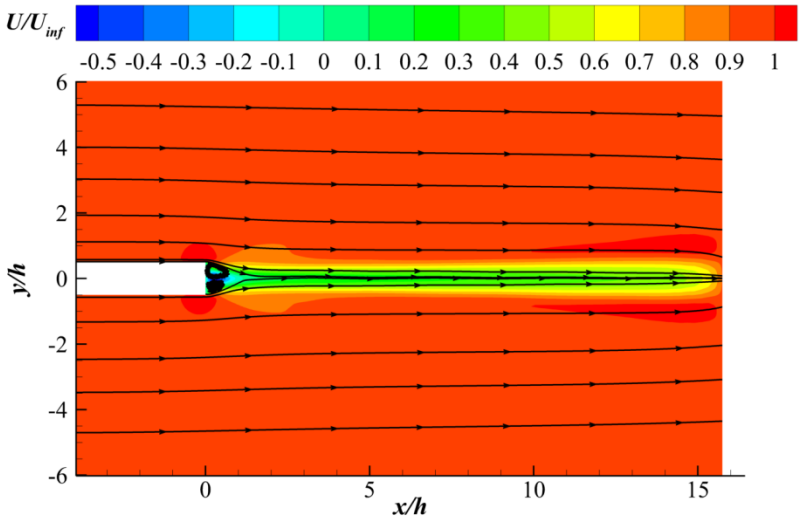


Fig. II.54: Average streamwise velocity field (Euler simulation of a blunt trailing edge wake)

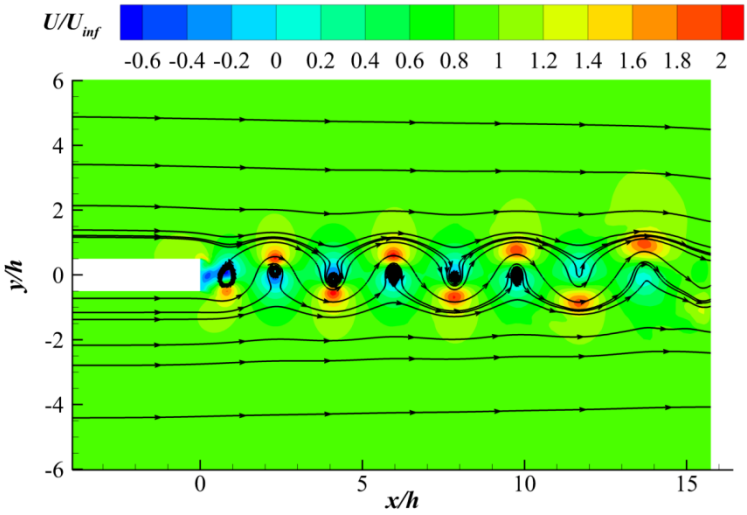


Fig. II.55: Instantaneous streamwise velocity field (Euler simulation of a blunt trailing edge wake)

4.2. Flow around a supercritical airfoil in transonic conditions

SE has been recently used to estimate the flow around an airfoil. For instance, Garcia-Sagrado et al. [91] studied the wake downstream a NACA0012 profile and Carabello et al. [92] the detached flow downstream a NACA0015 profile. SE was however never used with shocked flow (to the author's knowledge). Therefore, the use of the flow around an OAT15A airfoil in transonic conditions represents an interesting test case. The OAT15A is an airfoil with a chord c = 230 mm.

Moreover, this flow had already been simulated at the Onera. Thus, it was possible to use previously made databases or to rapidly run again the simulations to obtain additional data. Another interesting point was the availability of URANS and ZDES simulations which could be used to investigate the impact of turbulence on the SE performances.

4.2.1. Unsteady Reynolds Average Navier-Stokes (URANS) simulation

The simulation was run using the elsA software. The mesh is a 2D structured mesh of about 3×10^5 cells (see Fig. II.56). The space discretization is carried out using a 2nd order backward Euler scheme (with multigrid treatment). The Dual Time Step algorithm is used for the time integration with a time step of 1.7×10^{-5} s and a maximum number of sub-iteration of 100. The turbulence model is the k-ω Wilcox one. The aerodynamics conditions considered are 4.5° for the angle of attack, 3.10^6 for the Reynolds number (based on the chord Re_c) and 0.73 for the freestream Mach number. In these conditions the buffet frequency is around 75 Hz ($St \approx 0.066$).

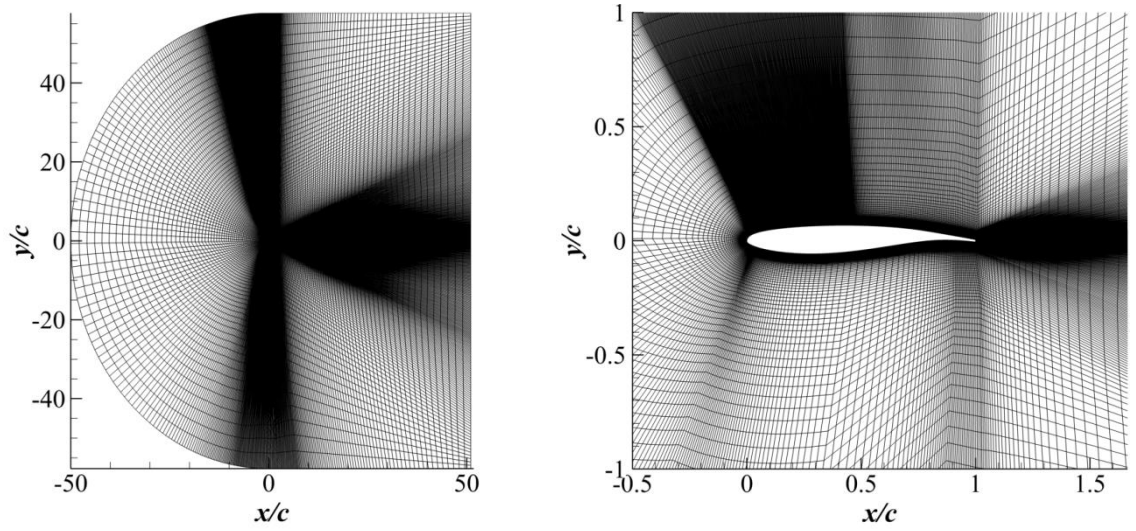
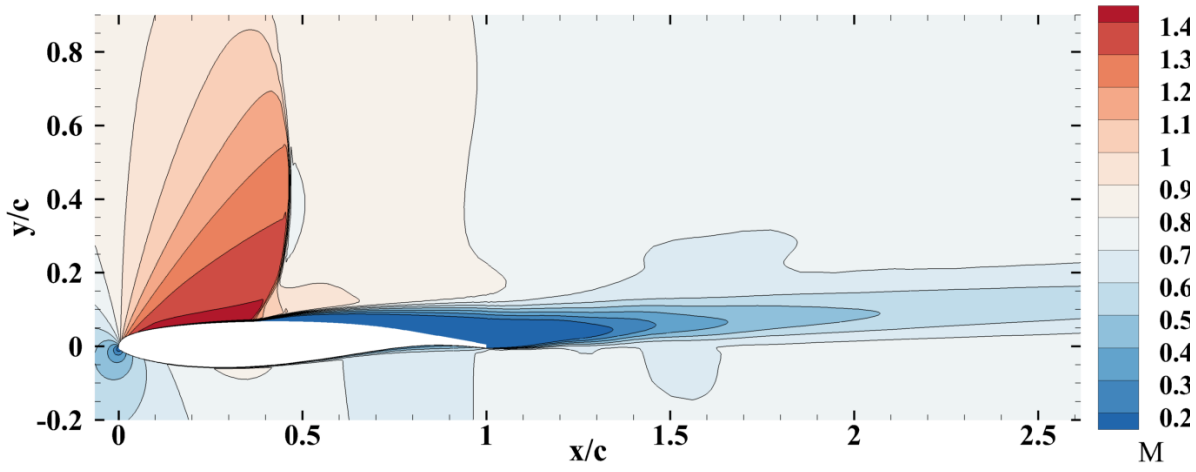


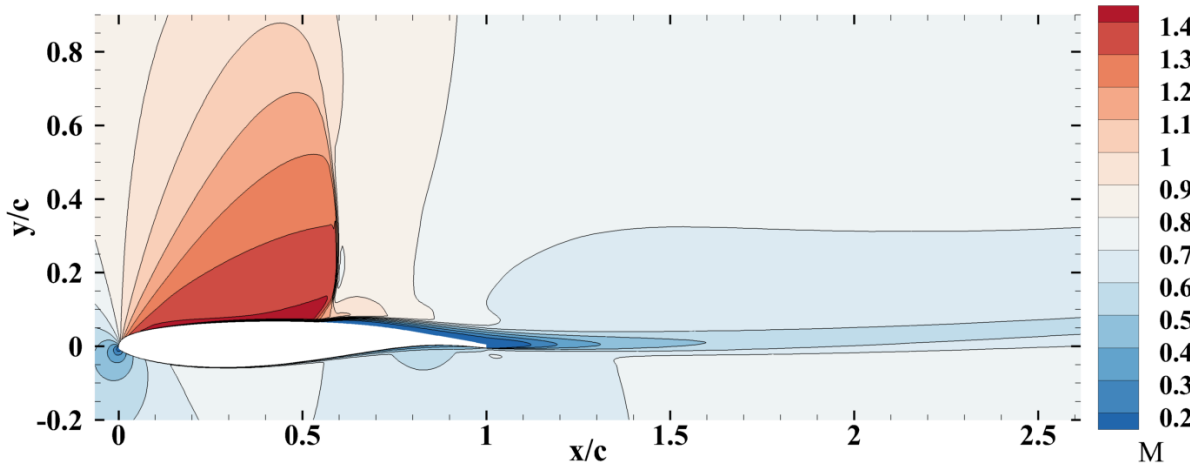
Fig. II.56: Views of the mesh used for the OAT15A URANS simulation

Fig. II.57 shows instantaneous flow fields at different time steps of the simulation. A shock is clearly identified on the suction side. Its oscillation is also highlighted between the 3 views. In addition, a separation is observed downstream the shock. The detachment occurs right downstream of the shock. Slight oscillations of the wake can be seen in Fig. II.57.c), but there is no clear vortex shedding.

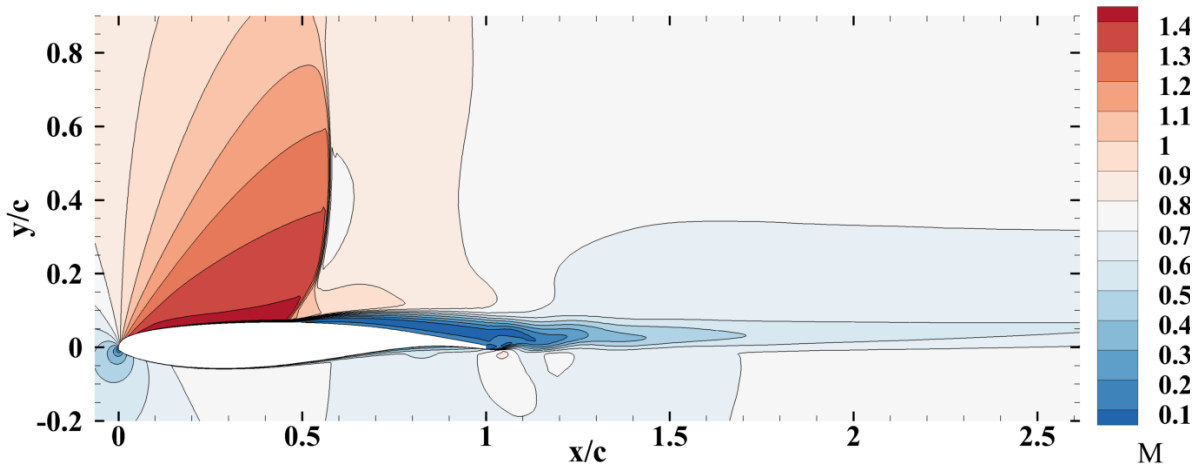
2 850 snapshots were extracted in order to use SE methods. They correspond to less than 4 periods of the shock oscillation (one period covering about 750 time steps). The two velocity components, the static pressure and the density were extracted in these snapshots. The first 1 350 snapshots are put aside in order to allow the use of MTD estimation techniques. Then the next 750 fields are used as a training set, and the last 750 as a validation set. Similarly to the Euler simulation, the URANS simulation result is not a stochastic process (the turbulence is completely modeled, not resolved). All buffet periods are thus identical and consequently the knowledge of one period is enough to compute the averages required by SE methods. It is considered that about 750 samples by buffet period are enough to get a good estimation of these averages. Also since the time step used for the simulation is not a multiple of the shock oscillation period, instantaneous fields from the training set and validation set are still a little bit different, allowing us to form a validation set different from the training set.



a)



b)



c)

Fig. II.57: Visualizations of the Mach number field. a) Shock at its most upstream position, b) shock at its most downstream position, c) shock at an intermediary position (OAT15A URANS)

4.2.2. Zonal Detached Eddy Simulation (ZDES)

4.2.2.1. Description of the simulation

A ZDES (Deck [93]) was also carried out using the elsA software. The mesh is a 3D structured mesh of approximately 2×10^6 cells. Views of the mesh are displayed in Fig. II.58. The simulation is run using the URANS equations in the black part of the mesh and ZDES mode 2 in the green part. The Navier-Stokes equations are discretized by means of a 2nd order accurate upwind finite volume scheme and a cell-centered discretization. The Euler fluxes are discretized using the AUSMP (Advection Upstream Splitting Method Preconditioned) upwind scheme. Time integration is carried out with a second-order Gear scheme with a time step of 5×10^{-7} s. The aerodynamics conditions considered for the ZDES are an angle of attack of 3.5°, a Reynolds number (based on the chord Re_c) of 3×10^6 and a freestream Mach number of 0.73.

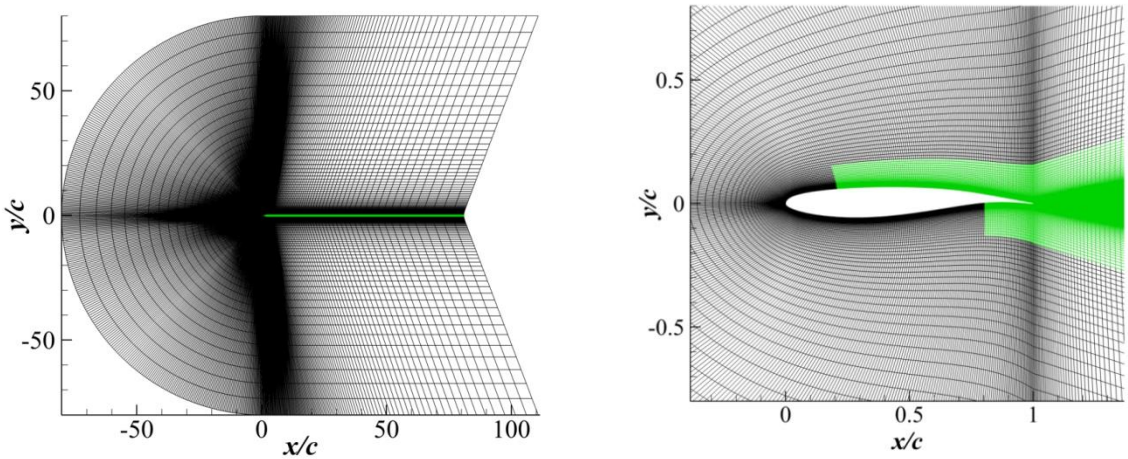


Fig. II.58: Views of the OAT15A ZDES mesh. ZDES mode 2 is used in green blocks. URANS equations are used in the other blocks.

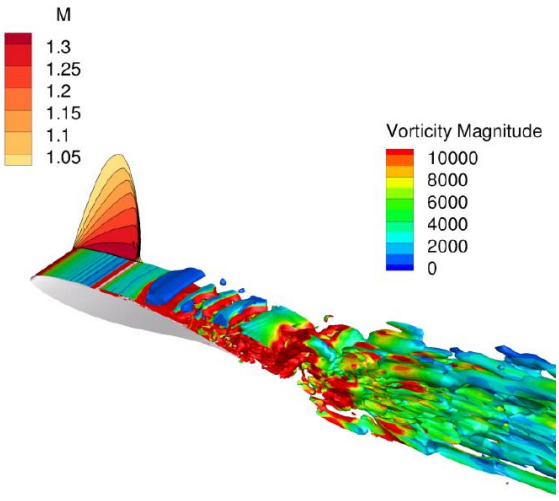
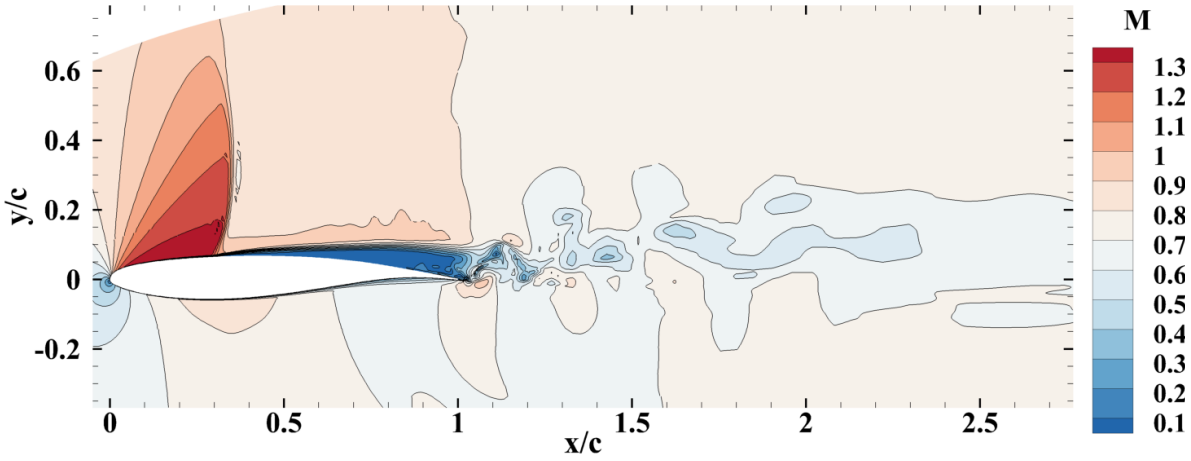


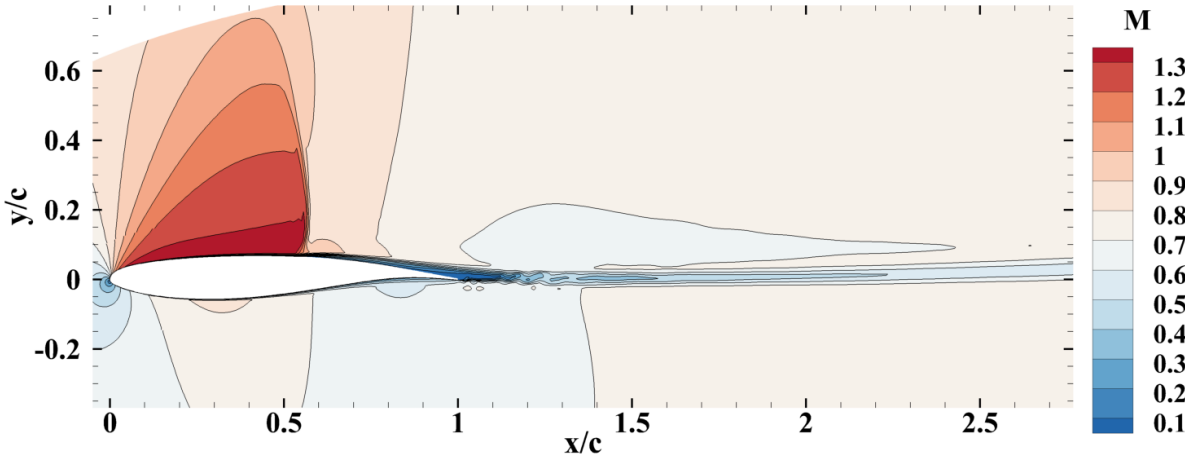
Fig. II.59: Isosurfaces of Q-criterion colored by the vorticity magnitude, and contours of Mach number in a transverse plane (Toubin [94])

The shock oscillation frequency is found around 70 Hz ($St \approx 0.062$) and covers approximately 29 000 time steps. About 10 periods of the buffet oscillation were computed (once the transitory

phase was over). As in the URANS case, the first period is put aside in order to be used with MTD estimation techniques. The following 8 periods are used as a training set and the last period as a validation set. Snapshots are extracted with a sampling frequency of 40 kHz. One oscillation period spans about 580 snapshots. They contain the three velocity components, the density and the static pressure.



a)



b)

Fig. II.60: Visualizations of the Mach number field in a spanwise plane. a) Shock at its most upstream position, b) shock at its most downstream position (OAT15A ZDES)

Visualizations of a slice the flow are given in Fig. II.60 when the shock is most upstream a) and downstream b). As for the URANS case, a shock formed on the suction side of the profile and the flow detached just downstream of this shock. But unlike the results of the URANS simulation, vortices are shed from the airfoil trailing edge (especially when the shock is the most upstream). Isosurfaces of Q-criterion colored by the vorticity magnitude and contours of the shock wave Mach number are displayed in Fig. II.59 when the shock is in an upstream position. The shedding of small and large scale structures in the wake can be observed.

4.2.2.2. Convergence of the statistics

Studying the convergence of statistics in the ZDES results is a complicated matter. Indeed, parts of the flow are simulated using URANS equations, while some parts are simulated using DES

method. In URANS parts, the process is not stochastic. In DES regions, solving part of the turbulence make the process stochastic. To determine the number of statistically independent samples, the autocorrelation of 5 pressure data and 8 velocity data are investigated. The locations of the several data points are shown in Fig. II.61. The autocorrelations of the pressure data are shown in Fig. II.62 and the autocorrelations of the streamwise velocity for the 8 selected locations in Fig. II.63. From these figures the five pressure data appear deterministic and some velocity data appear stochastic. Streamwise velocities at the first four locations are deterministic. Otherwise they appear to be stochastic. However, the signals at locations 6 and 7 present a strong general pattern and seem almost deterministic. An integral time scale is calculated only for the streamwise velocity signals 5 and 8 for which the autocorrelation falls and remains under 1/e. T_i is found, in both cases, to be approximately 1.5×10^{-4} s. Based on this integral time scale, there are about 800 independent samples in the training set formed of 4 640 snapshots.

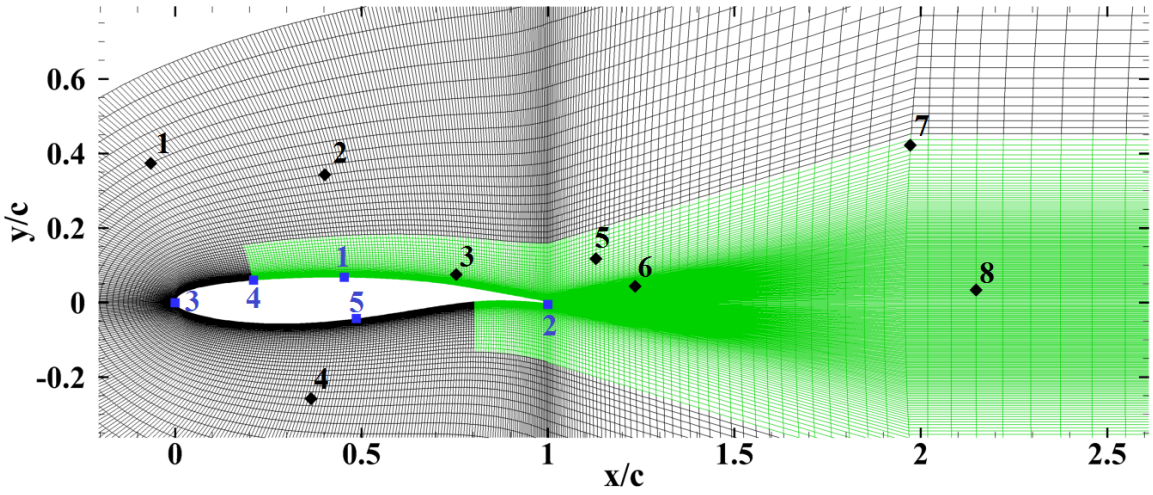


Fig. II.61: Positions of the pressure and velocity data used for the convergence study. Blue squares: pressure data, black squares: velocity data

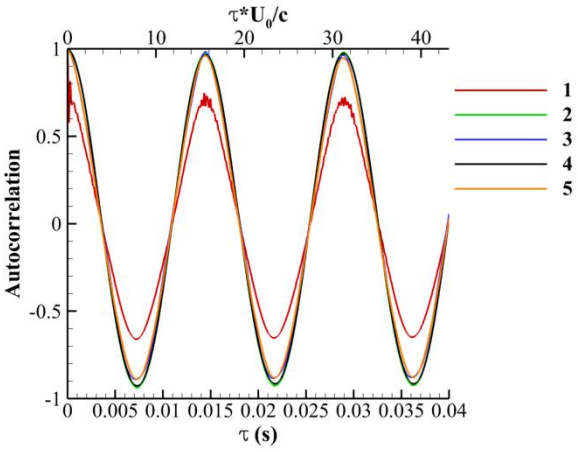


Fig. II.62: Autocorrelation as function of the delay for 5 pressure signals at the skin of the OAT15A airfoil

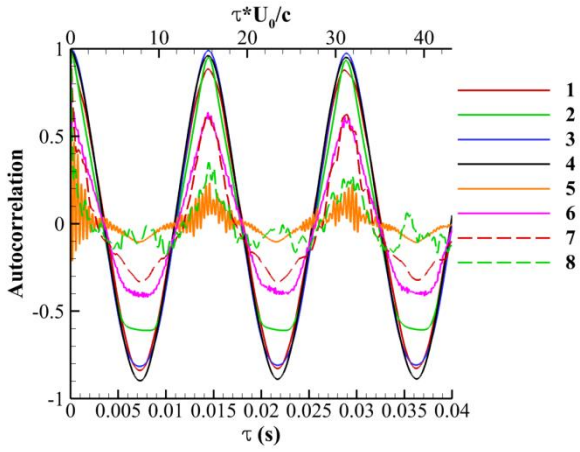


Fig. II.63: Autocorrelation as function of the delay for 8 streamwise velocity data

The 15 combinations of $\langle P_i P_j \rangle$ $(i \leq j)$ for the 5 pressure data obtained from the training set are displayed in Fig. II.64. Once more the error bars correspond to the 95% confidence interval. This interval is calculated considering 800 independent samples. The maximum relative uncertainty reaches 56% but all the remaining values are under 18% (in particular 12 of them are under 6%).

The convergence of the covariance $\langle P_i P_j \rangle$ seems satisfying. For the 3rd and 4th order moments, the relative uncertainties are much higher and these moments are badly converged. The values of $\langle u(x_i) P_j \rangle$ are plotted in Fig. II.65 with the errors bars corresponding to the 95% confidence interval (for the vertical velocity). The values of $\langle u(x_i) P_j \rangle$ are insufficiently converged. The relative uncertainties are also particularly high, with a large amount of values higher than 100%. The similar conclusions are reached for the $\langle u(x_i) P_j P_k \rangle$ moments. In overall, the statistics from the training set are not well converged. Nevertheless, in the following, the performances of the SE will be evaluated using validation set, thus it will be possible to directly evaluated if the model suffers from overfitting.

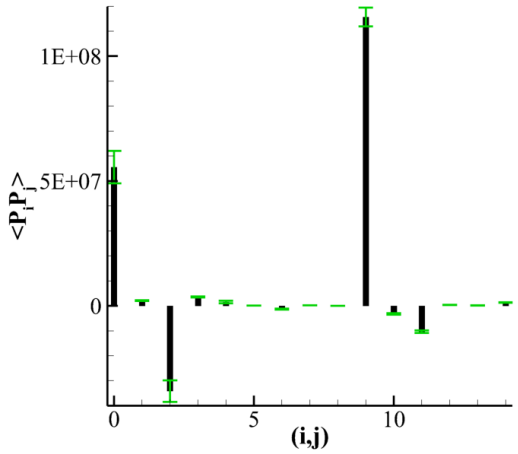


Fig. II.64: Covariance values between the pressure data. Error bars correspond to the 95% confidence interval.

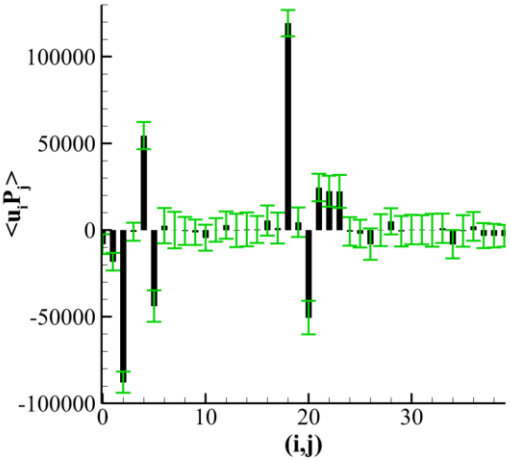


Fig. II.65: $\langle u(x_i) P_j \rangle$ values with their 95% confidence interval

4.3. 2D Backward facing step flow

4.3.1. Description of the simulation

The goal of this simulation is to reproduce the backward facing step experiment presented in Chapter II.3.2. To do so, a ZDES of the same BFS was performed, only for the $U_\infty = 30$ m.s^{-1} case. Once more the elsA software of the Onera was used to perform the simulation. The mesh is a 3D structured mesh (see Fig. II.66 and Fig. II.67). It contains approximately 18 million cells. The zone upstream the step is simulated using URANS equations (displayed in green in Fig. II.66). The zone downstream the step is simulated using ZDES mode 2 (displayed in red in Fig. II.66). The computational domain dimensions are 39 h in length (about 15 h upstream of the step and 24 h downstream), 4 h in width, 10 h in height (downstream of the step). The mesh satisfies $y^+ \leq 1$ for the boundary layer. In addition, at the step, the following characteristics are verified:

– $\Delta x = 62.5$ µm $\leq \delta\omega/2 \approx 1.25$ mm
– $\Delta y = 0.01$ mm $\leq \delta\omega/15 \approx 0.17$ mm
– $\Delta z = 0.75$ mm

Space discretization uses AUSMP (Advection Upstream Splitting Method Preconditioned) scheme with a 5th order MUSCL reconstruction. The time integration scheme is the Gear one with 7 sub-iterations. The time step is 0.4 µs in order to satisfy the CFL condition $CFL_{max} \leq 15$ in the mixing layer. The upper boundary condition is a non reflective condition. Lateral boundaries use periodicity conditions. Walls are treated as viscous walls (adiabatic conditions in elsA). A pressure

of 101 325 Pa is imposed on the downstream boundary condition. At last a velocity profile is imposed at the upstream boundary condition. Indeed, in order to match the experimental conditions, a velocity profile was chosen for this boundary condition, such as the boundary layer obtained in RANS simulation, 10 mm upstream the step, matches the experimental boundary layer at the same location.

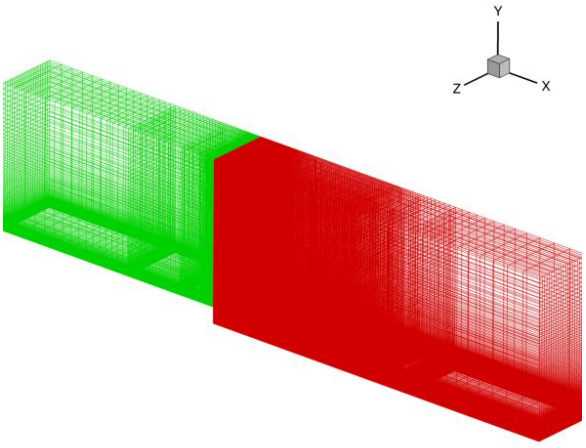


Fig. II.66: 3D view of BFS ZDES mesh (green zone is simulated using URANS equations, red zone using ZDES mode 2)

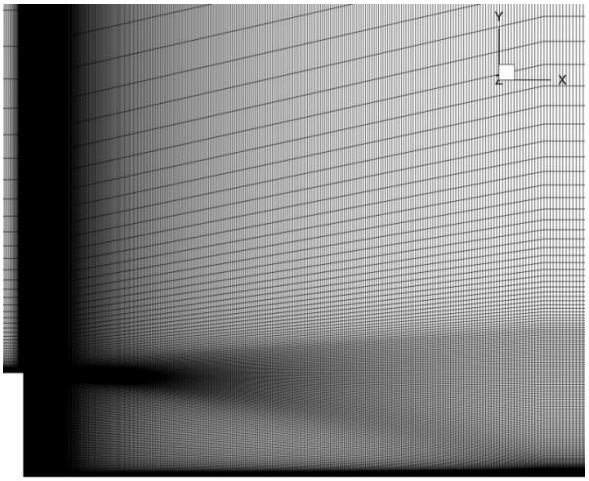


Fig. II.67: Longitudinal cut of the mesh

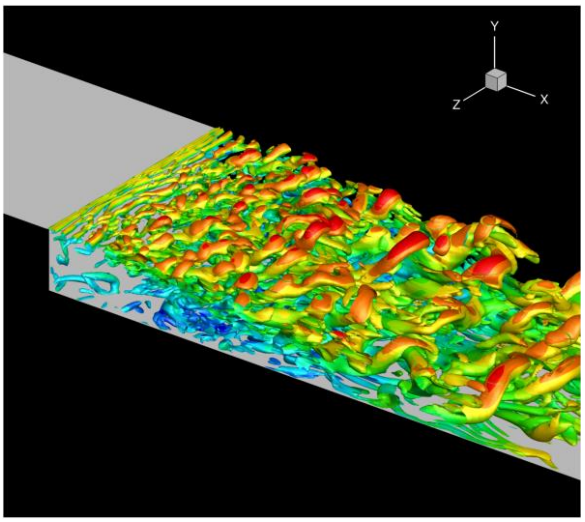


Fig. II.68: Isosurfaces of Q-criterion colored by the streamwise velocity

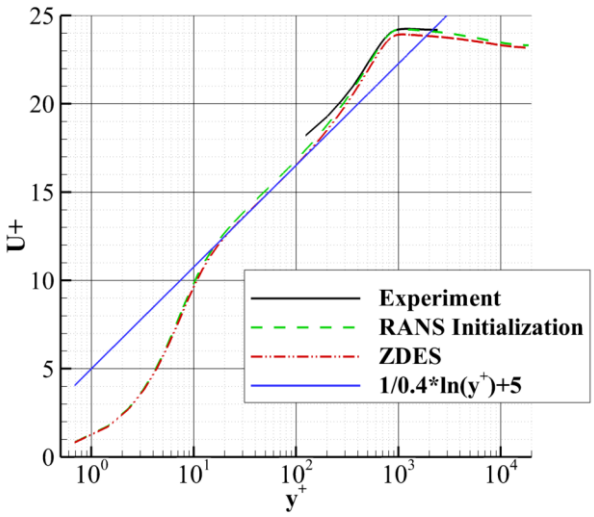


Fig. II.69: Boundary layer profiles 10 mm upstream of the step. Comparison between PIV, RANS and ZDES results

About 500 ms have been simulated, but the transitory phase lasts for about 240 ms, leaving the last 260 ms to be used. This represents about 26 shedding periods. Isosurfaces of the Q-criterion, coloured by the streamwise velocity, are shown in Fig. II.68. To validate the simulation, its results are compared with the experimental ones. First the boundary layer profiles 10 mm upstream the step obtained in the experiment, in the RANS initialization of ZDES and in the ZDES are compared in Fig. II.69. The agreement between the 3 profiles is good. The reattachment location is found around 6.3 *h* which is consistent with the experiment value (5.9 *h*). Time-averaged streamwise velocity profiles from the experiment and the ZDES are also compared in Fig. II.70. The streamwise

and vertical root-mean-squared velocity profiles are compared between the experiment and the ZDES in Fig. II.71 and Fig. II.72. ZDES time-averaged streamwise velocity profiles exhibit an excellent agreement with the experimental ones. The root-mean-squared velocity quantities are also in very good agreement. One can clearly observe the absence of turbulence in the boundary layer upstream of the step in the simulated case, which is due to the RANS simulation in this zone. For the longitudinal quantity, simulated values are slightly underestimated whereas, for the vertical quantity, they are slightly overestimated. In overall, the agreement between the ZDES and the experiment is completely satisfying.

The numerical database is composed of the longitudinal plane where the density, the 3 velocity components and the static pressure are extracted. Initially, the frequency of sampling was set to 20 kHz but was then increased at 250 kHz to obtain more samples (even if they were correlated). Thus the database contains 2 300 snapshots at 20 kHz and 35 900 snapshots at 250 kHz.

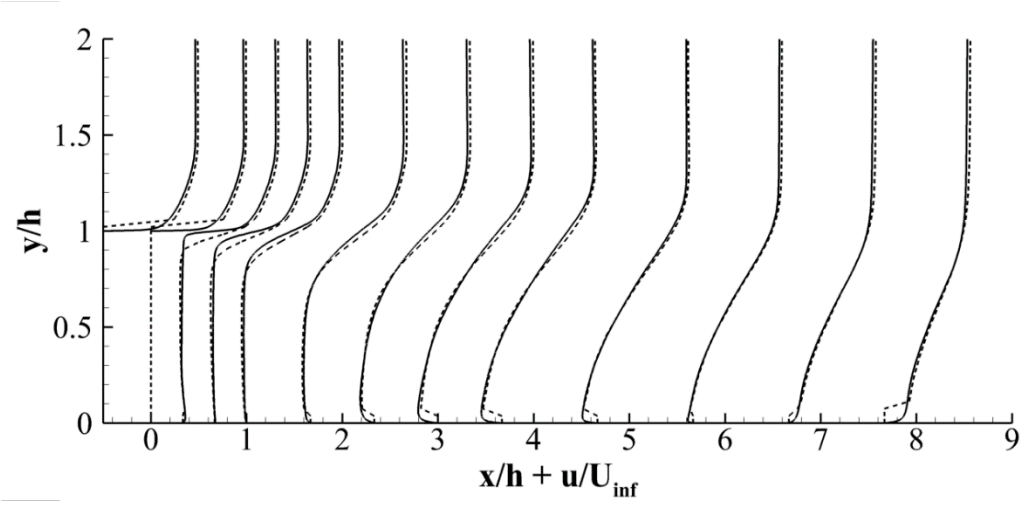


Fig. II.70: Comparison of normalized average streamwise velocity profiles (dashed lines: PIV results; solid lines: ZDES results)

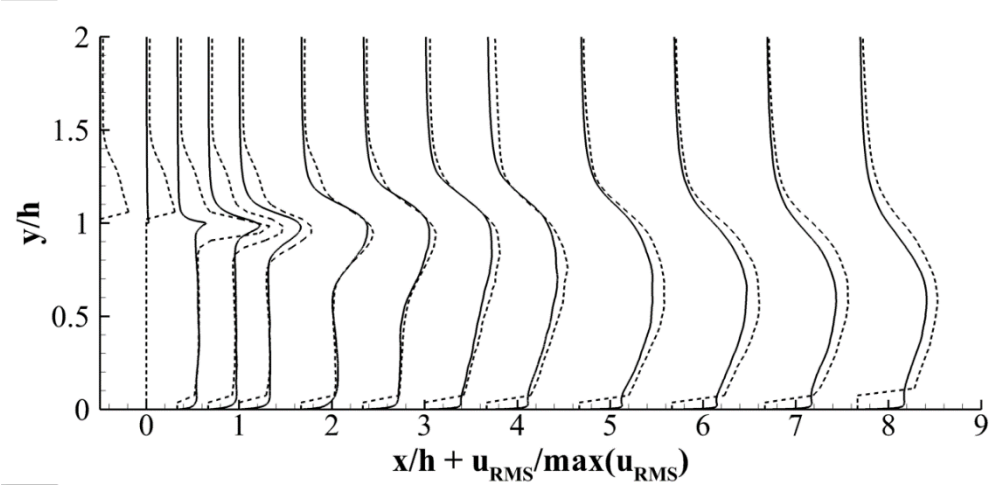


Fig. II.71: Normalized streamwise root-mean-squared velocity profiles (dashed lines: PIV results; solid lines: ZDES results)

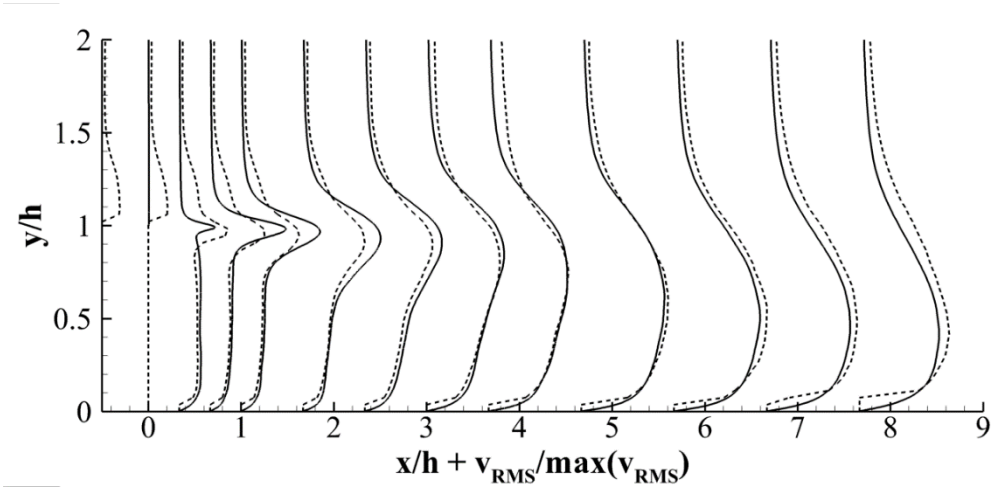


Fig. II.72: Normalized vertical root-mean-squared velocity profiles (dashed lines: PIV results; solid lines: ZDES results)

In addition, unsteady features of the simulation present a good agreement with the literature. The PSD computed at point (x = 3 h, y = 1.5 h) is shown in Fig. II.73. A peak is observed around 300 Hz (St = 0.3 based on the step height). This frequency corresponds to the Kelvin Helmholtz instability as expected from the literature (Chun et al. [95]). The spectrum at (x = 7 h, y = 0.1 h) (near the reattachment) is displayed in Fig. II.74. The peak is about 120 Hz (St = 0.12) which is consistent with the expected shedding frequency.

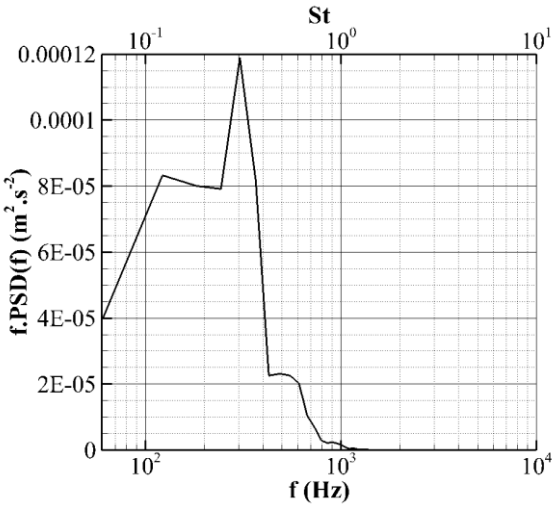


Fig. II.73: PSD of the streamwise velocity at (x = 3 h, y = 1.5 h)

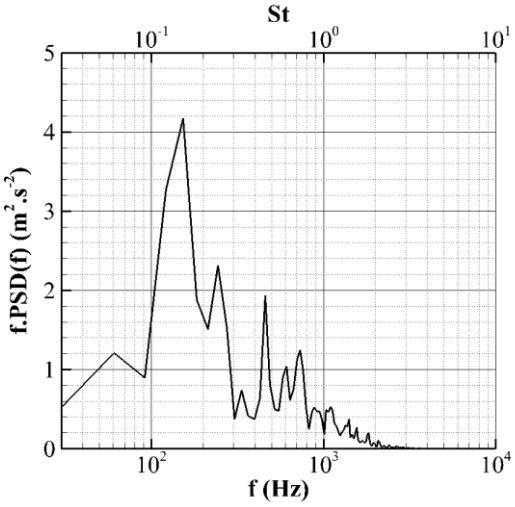


Fig. II.74: PSD of the streamwise velocity at (x = 7 h, y = 0.1 h)

4.3.2. Convergence of statistics

Once more, the integral time scales are used to get an estimate of the number of independent samples available. Thus, the autocorrelation functions are first calculated for the 17 pressure sensors at the same location as in the experiments (only data sampled at 250 kHz are used). They are plotted in Fig. II.75. As one can observe, the autocorrelation is not well converged. Indeed, the functions are far from being as smooth as in Fig. II.16. A strange behavior is obtained for the sensor upstream the step and is likely due to the fact that the area is simulated using URANS equations. Nevertheless, from the third sensor downstream the step, the behavior of the autocorrelation function is closer to the one observed in the corresponding experiments. The integral time scales are found between 0.5

ms and 2.5 ms, which is slightly lower than values obtained with the experimental data. Integral time scales of the 13 velocity data points (defined in Chapter II.3.2.4.1, see Fig. II.20) are of the same order of magnitude (the autocorrelation functions for the streamwise velocity are shown in Fig. II.76 for illustration). Since the experimental data are better converged, it is decided to use the integral time scale experimentally estimated to determine the number of independent samples. In addition, since the experimental values are higher, the constraint on the convergence estimation is thus stronger. Therefore, based on an integral time scale of 5 ms, the 2 300 snapshots sampled at 20 kHz represents about 23 independent samples and the 35 900 snapshots taken at 250 kHz about 29 independent samples. Thus the number of independent samples is extremely limited and the statistics are not converged.

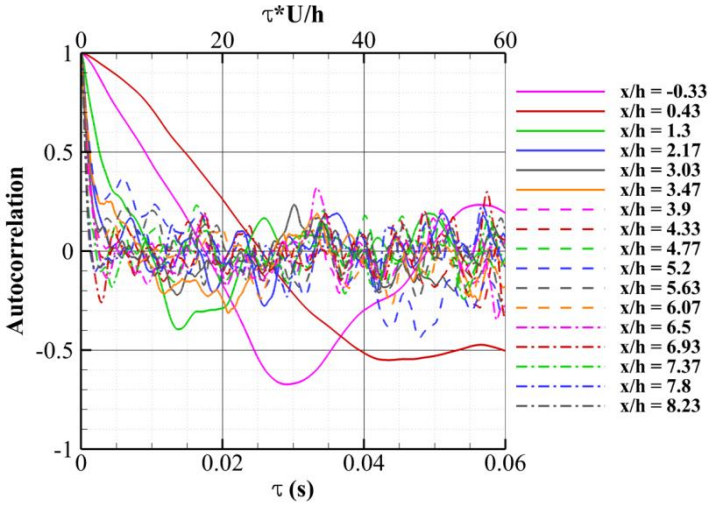


Fig. II.75: Autocorrelation as function of the delay for the 17 pressure sensors (extracted from the BFS ZDES)

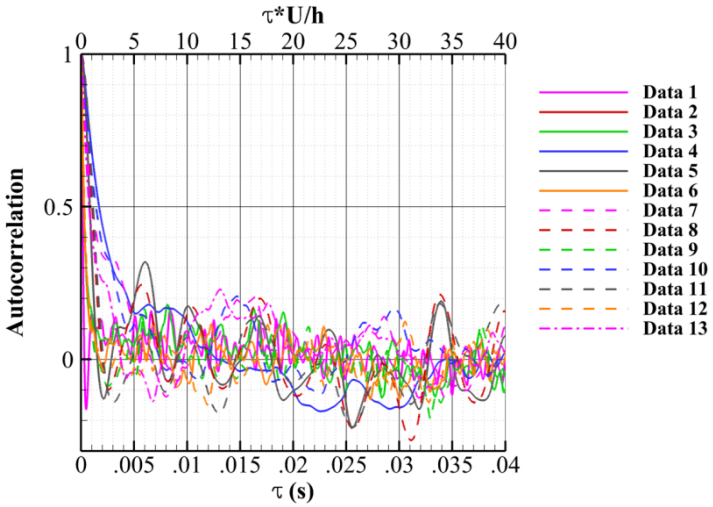


Fig. II.76: Autocorrelation as function of the delay for the 13 streamwise velocity data extracted from BFS ZDES field

5. Chapter summary

The mathematical framework of the Stochastic Estimation and its extensions has been first detailed. In relation to the literature and SE principles, several test cases have been chosen in order to test the different SE methods.

Several experimental and numerical databases are therefore at our disposal and have been described. A very basic simulation of the wake downstream of a blunt trailing edge can be used to validate the implementation of the SE methods. Then, numerical and experimental databases of a 2D backward facing step flow have been generated. This geometry corresponds to a flow for which SE has already been applied in the literature, but the present databases have been obtained at a higher Reynolds number. As a second step toward the final goal of the HYBEXCIT project (flow around the helideck of a frigate), databases of the flow downstream of a wall mounted cube was generated experimentally. At last, in order to test the SE on an uncommon configuration, simulations of the flow around an OAT15A airfoil in transonic conditions (thus a shocked flow) have been performed.

For all databases, special care has been given to the convergence of the statistical moments used by the SE. For the experimental databases, the number of independent samples appears to be insufficient to converge moments of order 3 or 4 (whatever the case). For the low speed PIV campaigns, order 2 moments seem converged. When using high speed PIV, only in the cube case, the order 2 moments seem to be sufficiently converged. For the numerical databases, one has to truly worry for the convergence where the turbulence is not completely modeled. For the ZDES of the OAT15A airfoil in transonic conditions, only the covariance between the pressure sensors can be considered to be correctly converged. For the ZDES of the BFS, no statistical moments are converged. In the following work, the SE methods will be tested on validation sets, thus the impact of the convergence will be directly appreciated from the difference between the results obtained on the training set and on the validation set.

Chapter III. Performance evaluation of the Stochastic Estimation for the reconstruction and prediction of turbulent flows

The aim of this chapter is to compare the abilities of the SE and its extensions to reconstruct and predict turbulent flows of different complexity. For incompressible flow, the specification of the three components of velocity is enough to uniquely define the flow. In this work, except for the OAT15A airfoil, flows can be considered as being incompressible due to the relative low Mach number. Therefore, the study will be limited to the reconstruction and prediction of velocity fields. In addition, for applications such as the landing of a helicopter on a frigate, flow velocity information is of prime importance.

To evaluate the performance of the SE and its extensions, these techniques are applied to flows described in Chapter II. First, SE is applied to a blunt trailing edge wake, mainly to validate the implementation of the method. Then, it is used for the estimation of the OAT15A airfoil flow in transonic conditions from the ZDES simulation, of the backward facing step flow and of the wall mounted cube flow. In all these test cases, the quality of the reconstruction and prediction will be assessed using the determination coefficient as a global quality criterion but also in term of normalized mean square error for local evaluation.

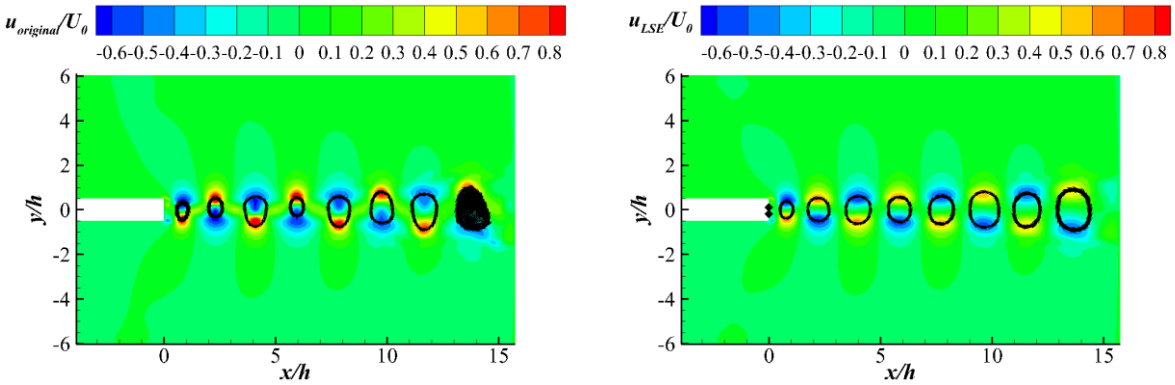
1. Application to the wake downstream of a blunt trailing edge

1.1. Single-time Stochastic Estimation

1.1.1. Linear Stochastic Estimation (LSE) and Quadratic Stochastic Estimation (QSE)

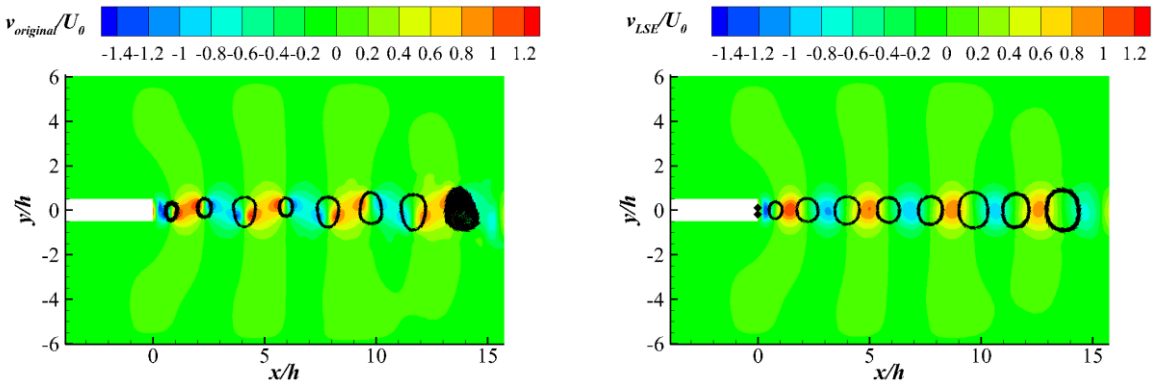
LSE is used to estimate the velocity field in the wake downstream of a blunt trailing edge simulated using Euler equations. First, only two pressure data extracted from the simulation at the blunt trailing edge are used as conditional events (their location is shown in Fig. III.1.b) using black diamonds). An example of an instantaneous field (from the validation set) and its estimation is shown in Fig. III.1 for the streamwise velocity and in Fig. III.2 for the vertical velocity. As expected from results in the literature, the estimation is good. The von Kármán vortex street is well reproduced, with vortices at the correct position. The determination coefficient R^2 is about 80%. Maps of the normalized mean square error for the streamwise and vertical velocity prediction (validation set) are plotted in Fig. III.3. The error is mainly located at the center of the vortex street for the streamwise velocity. For the vertical velocity, the error is spread on both upper and lower side of the vortex street. The average correlation module (see equation (II.29)) between the two pressure sensors and the streamwise and vertical velocity fields is shown in Fig. III.4. The structures depicted by the low values of the average correlation module in both cases are very similar in shape with the ones from the highest values of the normalized mean square error. This observation clearly highlights the relationship between estimation quality and the correlation between the sensors and the field to estimate. It is thus possible to deduce, directly from the training set, areas where LSE will likely fail to predict the flow field (outside the training set) by looking at the average correlation module.

The quality of integral quantities estimation has also been evaluated. The streamwise and vertical root-mean-squared velocities are displayed in Fig. III.5 and Fig. III.6. Levels and general trends of both quantities are well estimated by the LSE with only 2 pressure sensors.



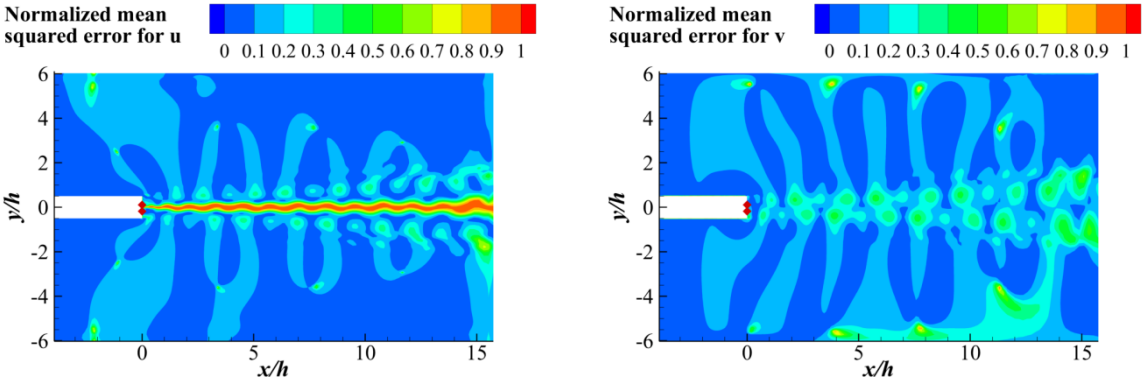
a) b)

Fig. III.1: Comparison between a) an original instantaneous streamwise velocity field (from the validation set) and b) its reconstruction by LSE using 2 pressure sensors (black diamonds).



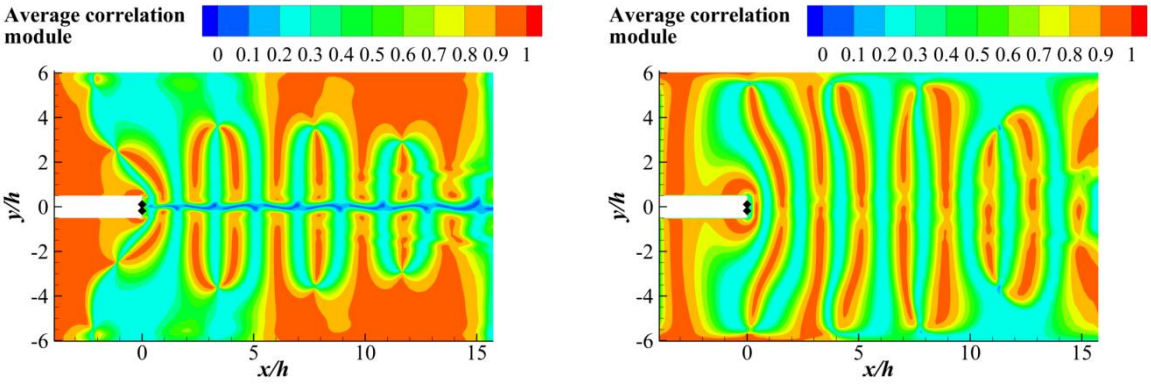
a) b)

Fig. III.2: Comparison between a) an original instantaneous vertical velocity field (from the validation set) and b) its reconstruction by LSE using 2 pressure sensors (black diamonds).



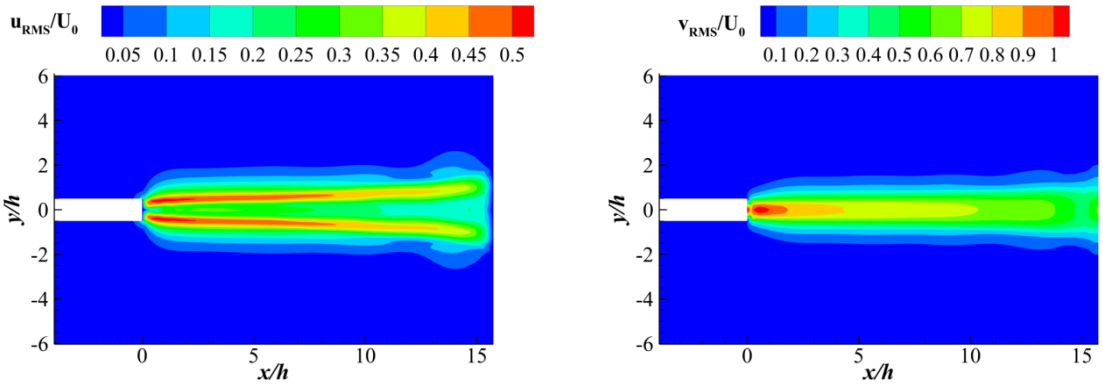
a) b)

Fig. III.3: Normalized mean square error of the validation set for the streamwise velocity (a) and the vertical velocity (b)



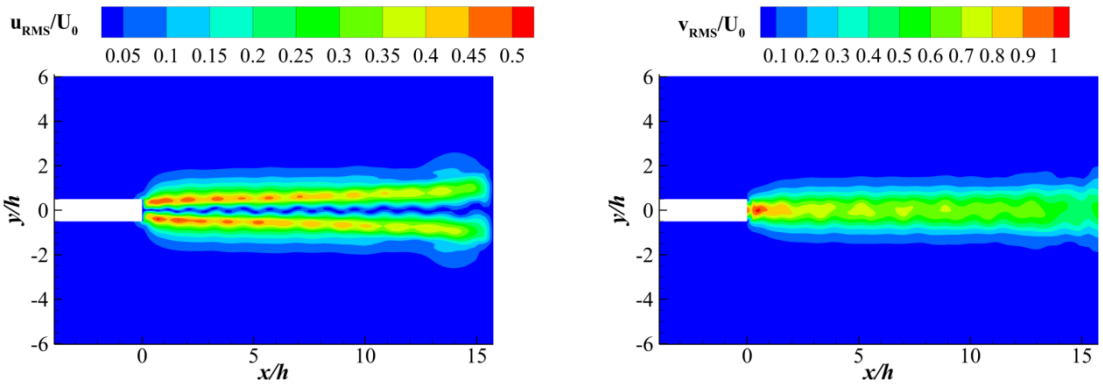
a) b)

Fig. III.4: Average correlation module between the 2 pressure sensors and the streamwise velocity field (a) and vertical velocity field (b)



a) b)

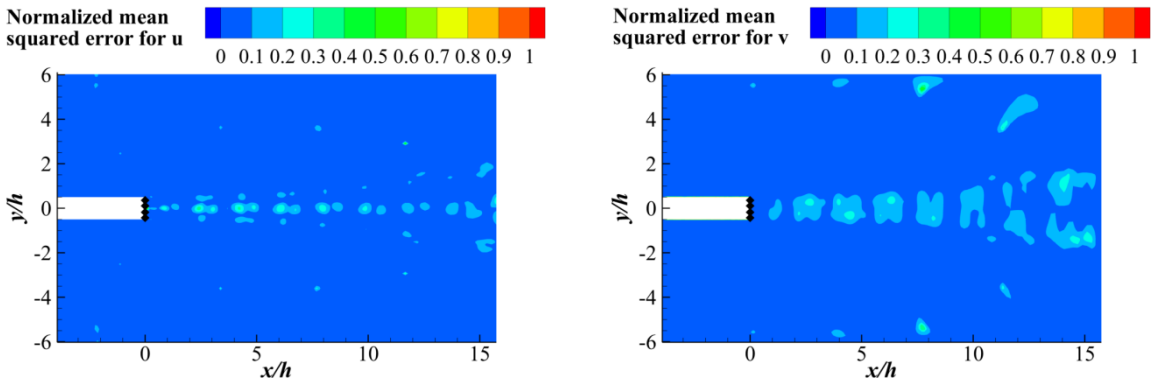
Fig. III.5: Original a) Streamwise and b) vertical root-mean-squared velocities



a) b)

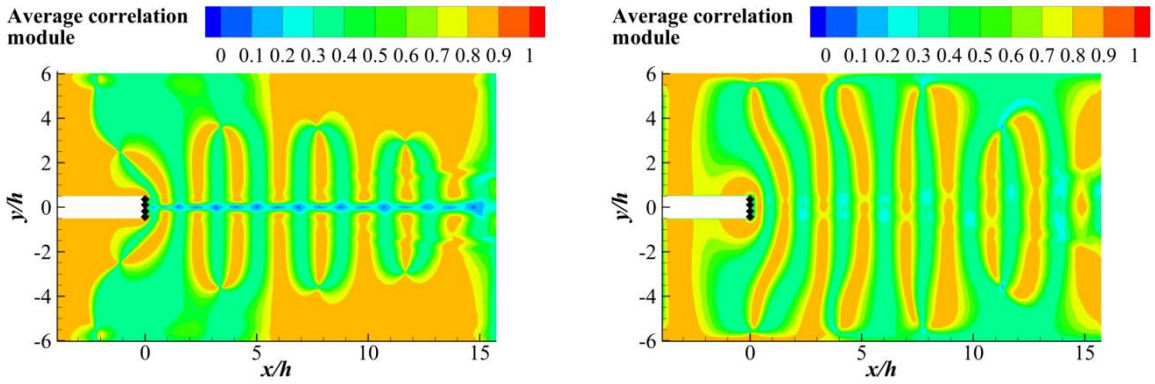
Fig. III.6: a) Streamwise and b) vertical root-mean-square velocities estimated by LSE with 2 pressure sensors (validation set)

The impact of adding more information by using more sensors has also been investigated. Fig. III.7 shows the normalized mean square error (of the validation set) for a LSE using 4 pressure sensors. The error is greatly reduced and the determination coefficient reaches 92%. As expected, the addition of conditional events on the prediction ability of the LSE is important. By looking at the maps of average correlation module (see Fig. III.8), it is still possible to link the region of highest error with the ones of lowest correlation. The overall effect of the addition of pressure sensors is here to better “distribute” the correlation. Indeed, the levels of average correlation module are less scattered in the case using 4 sensors than in the case using 2 sensors. The addition of sensors has also for effect to increase the VIF which goes from 3.3 to 10.2, however no overfitting is observed. This is due to the fact that the VIF remains relatively low in both cases and also to the small differences between the training and validation sets. The estimation of root-mean-square velocities is also improved by the use of additional sensors. The predicted root-mean-square velocities are in excellent agreement with the original ones (see Fig. H.1 in Annex H.1).



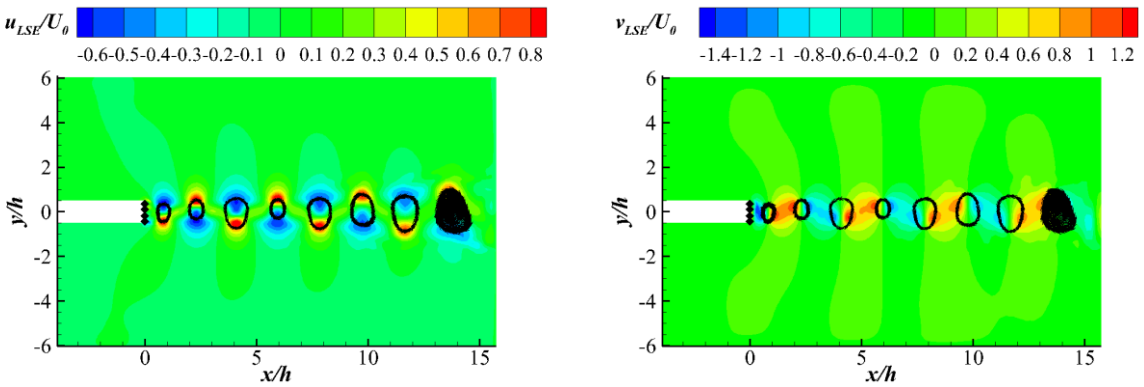
a) b)

Fig. III.7: Normalized mean square error of the validation set for the streamwise velocity (a) and the vertical velocity (b)



a) b)

Fig. III.8: Average correlation module between the 4 pressure sensors and the streamwise velocity field (a) or vertical velocity field (b)



a) b)

Fig. III.9: Estimation by QSE with 4 pressure sensors of an instantaneous velocity field a) streamwise velocity and b) vertical velocity

The use of QSE was then investigated. Once more, 2 and 4 pressure sensors are used as conditional events. The estimation of the same instantaneous velocity field as in Fig. III.1 and Fig. III.2 using QSE and 4 sensors is displayed in Fig. III.9. The estimated flow field in this figure is almost

indistinguishable from the original flow field. Both reconstruction (training data set) and prediction (validation) are excellent. QSE clearly improves the estimation accuracy compared to LSE. In terms of R^2, it reaches 92% for the QSE with 2 pressure sensors and more than 99.5% for the QSE with 4 pressure sensors.

1.1.2. Modified Stochastic Estimation (SE-POD)

The POD is computed using the 465 snapshots of the training set only. Fig. III.10 shows the energy spectrum of the POD decomposition. The first two POD modes contribute to more than 86% of the total TKE (each for approximatively the same amount). Each POD mode of higher order contributes to less than 4%. 7 POD modes are required to retain 99% of the TKE.

Contours of streamwise and vertical components of the POD modes 1 and 2 are plotted in Fig. III.11 and Fig. III.12. The spatial structures of these two POD modes are consistent with results found in the literature (Durgesh et al. [20], Tu et al. [65], Clark et al. [23]). The superposition of these two POD modes represents the dominant features of the von Kármán vortex shedding. The remaining POD modes correspond to smaller spatial scales and higher frequencies (than the vortex shedding) dynamics.

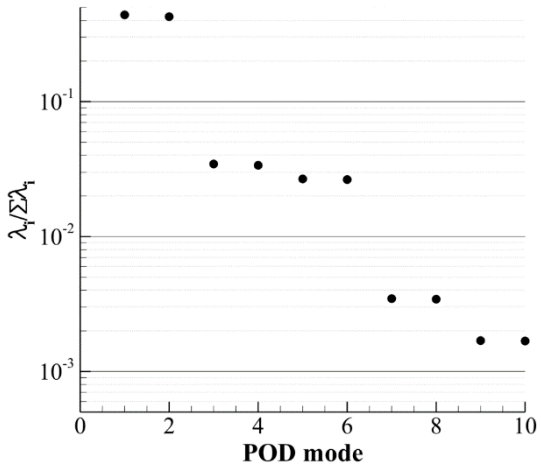
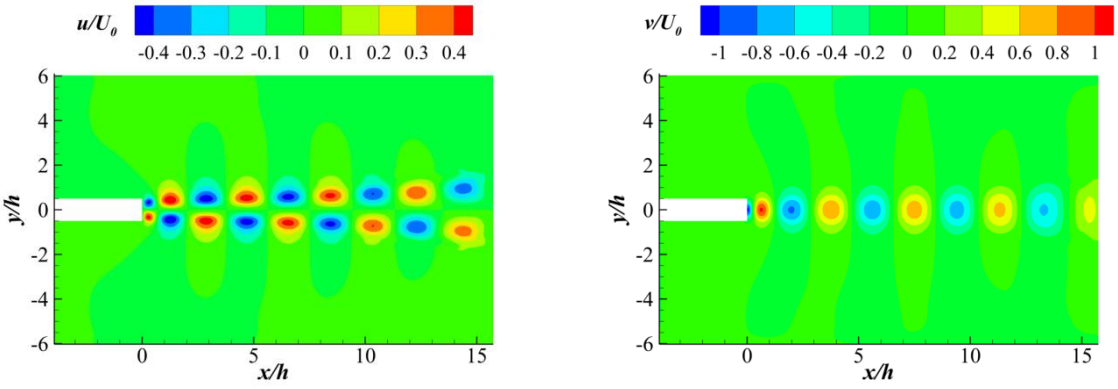
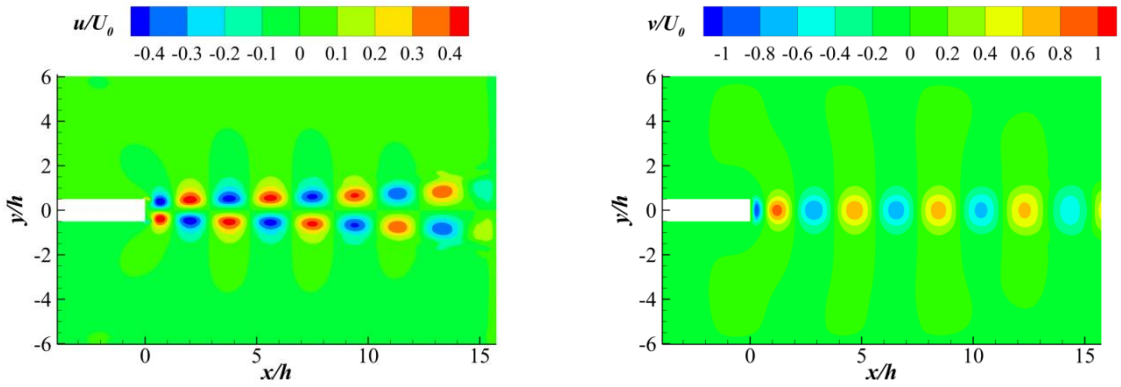


Fig. III.10: Energy spectrum of the first ten POD modes



a) b)

Fig. III.11: Contours of the streamwise (a) and vertical (b) components of the first POD mode



a) b)

Fig. III.12: Contours of the streamwise (a) and vertical (b) components of the second POD mode

LSE-POD using 2 and 4 sensors as previously is performed in which only the first seven POD modes are considered. The determination coefficient calculated with the original field (containing 100% of the TKE) is 80% using 2 sensors and 92% using 4 sensors. In both cases, R^2 is not improved using LSE-POD instead of LSE. The problem dimensions are however smaller since only 7 POD modes are required to obtain the same estimation as with the LSE. Only the 7 temporal POD coefficients are estimated instead of the 31 000 data of the entire flow field. Determination coefficients calculated for each POD mode separately ($R^2_{\text{POD Single}}$) are displayed in Fig. III.13 for both situations. In the case where 2 sensors are used, only the first 2 POD modes are well predicted ($R^2_{\text{POD Single}} > 85\%$). When 4 sensors are utilized, the first 2 POD modes are better predicted than in the 2 sensors case, but the prediction accuracy is much more improved for POD modes 3 and 4 ($R^2_{\text{POD Single}} > 85\%$ for both). In both cases, POD modes of rank 5 of higher are not correctly predicted ($R^2_{\text{POD Single}} < 10\%$). Therefore, the LSE appears to be able to predict (and reconstruct) only the first POD modes with high fidelity. For this flow, since most of the TKE is contained in the first 2 POD modes the overall estimation is excellent. However, if one is interested in the smaller and less energetic structures of the flow, more sensors are probably necessary.

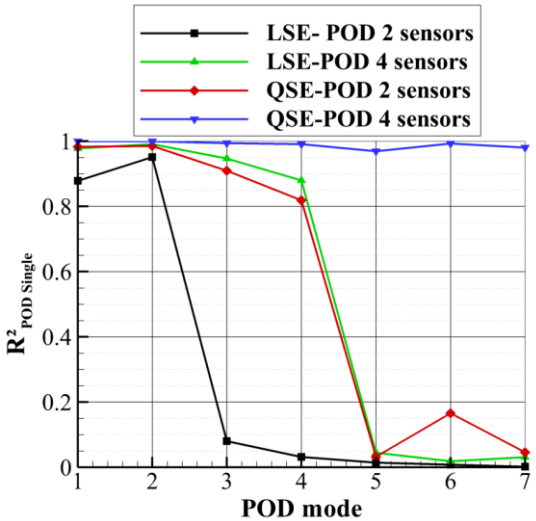


Fig. III.13: Determination coefficients calculated for each POD mode separately on the validation set (blunt trailing edge)

In agreement with previous observations, the QSE-POD reconstruction and prediction are largely improved compared to the LSE-POD. The same determination coefficients as in the case where no POD is used are achieved. $R^2_{\text{POD Single}}$ obtained on the validation set with QSE-POD are shown in Fig. III.13. As one can observe, with only 2 sensors, the QSE is able to match the accuracy of the LSE with 4 sensors. Using 4 sensors, the QSE outperforms the LSE with a very accurate reconstruction and prediction of the first 7 POD modes ($R^2_{\text{POD Single}} > 95\%$).

1.2. Multi-Time-Delay Stochastic Estimation (MTD-SE)

The use of MTD-SE was quickly investigated without any effort to optimize the number of delays. Indeed, the results obtained were already excellent. In this test case, the MTD-SE allows to reach a determination coefficient of 100% on the validation set. For instance, such performances are obtained with 2 sensors and 40 delays taken every 0.125 ms (thus 80 conditional events in total are used). In this situation, the first 7 POD modes are perfectly reconstructed and predicted. The MTD-SE clearly appears to improve the estimation quality. However, it is important to remind that in the present test case, the data to estimate present no stochastic behavior (there is no turbulence) and the validation set and training set are nearly identical.

1.3. Summary

Determination coefficients obtained on the validation set for the several situations tested are summarized in Table III.1. These results highlight the good performances of all techniques on a very simple flow, with a dynamic that is composed only of the von Kármán vortex street. The impact of adding sensors is also observed with an improvement in accuracy when 4 sensors are used instead of only 2. The use of POD does not improve the reconstruction of the initial flow field, but allows working on a smaller problem (the estimation of a small number of POD coefficients). QSE is shown to better perform than LSE (using the same amount of sensors). At last, the addition of delays also improves the reconstruction and prediction.

	2 sensors	4 sensors
LSE	80	92
LSE-POD (99% TKE)	80	92
QSE	92	99.5
QSE-POD (99% TKE)	92	99.5
MTD-LSE (40 delays)	100	100

Table III.1: Determination coefficients (expressed in %) comparison (wake behind a blunt trailing edge test case)

2. OAT15A airfoil in transonic conditions

In this part, only ZDES results are considered. Indeed, they correspond more closely to a real flow. The estimation of URANS velocity fields will be used in the next chapter to highlight some characteristics of the SE related to turbulence.

2.1. Single-time Stochastic Estimation

2.1.1. Linear Stochastic Estimation (LSE) and Quadratic Stochastic Estimation (QSE)

As previously, single-time LSE is first considered using wall pressure measurements (pressure data extracted from the simulation at the wall of the airfoil). Two sets of sensors which locations are chosen intuitively, based on the flow characteristics, are used. The two main phenomena are the shock oscillation and the vortex shedding in the wake. Thus, it seems logical to put a sensor near the mean shock location to "capture" the shock oscillation and to choose a sensor at the trailing edge to determine the wake flow. A sensor is also picked at the leading edge since it is justified to think that the flow around the airfoil is impacted by the physics at this location.

A set of 5 sensors is also formed using the 3 sensors of the previous set as well as one sensor on the pressure side (at about half the chord) and one sensor on the suction side between the leading edge and the mean shock location. The sensor locations are detailed in Table III.2. The set containing 3 sensors is referred to as "Intuitive P 3" and the set of 5 sensors as "Intuitive P 5". Sensors are shown with black squares in figures.

Intuitive P 3	Intuitive P 5
0.45 (s)	0.45 (s)
1 (s)	1 (s)
0 (s)	0 (s)
	0.21 (s)
	0.49 (p)

Table III.2: Pressure sensor positions (in *x*/*c*) for the two sets used for the estimation of the flow around the OAT15A airfoil. "s" stands for the suction side and "p" for the pressure side.

Since it was shown that the average correlation module could hold information on where the LSE will badly reconstruct and predict the field, we first displayed this data in Fig. III.14. From this figure, the wake region appears to be weakly correlated with both sets of sensors (whatever the velocity component considered). Thus, the LSE estimations should be of poor accuracy on this area. The streamwise velocity is also weakly correlated with both sets of sensors upstream and downstream the mean location of the shock. Thus high errors are also expected in these regions.

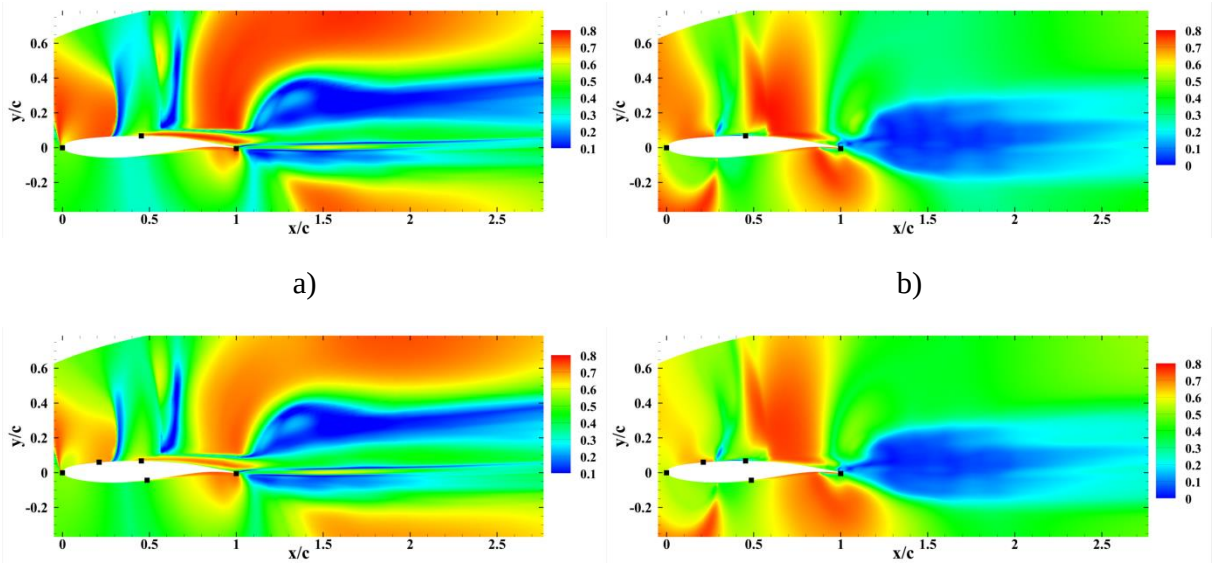


Fig. III.14: Average correlation module between the streamwise velocity (a), c)) or the vertical velocity (b), d)) and the set of pressure sensors (a), b) "intuitive P 3" set; c), d) "intuitive P 5" set)

An example of an instantaneous field (from the validation set) and its estimation is shown in Fig. III.15 (streamwise and vertical components are displayed). In the original field, the shock and vortex shedding can be observed. Concerning the estimation of the streamwise component of the velocity, the shock is somehow reproduced at the correct position in both 3 and 5 sensors cases (even if the levels of fluctuating streamwise velocity are not correct). But vortices in the wake are clearly not reproduced and are mostly rubbed out. The estimated wake seems to have been smoothed. In addition, the levels of the estimated fluctuations are lower than the original ones. These levels are however higher when using 5 sensors than when using 3 sensors. Similar observations can be done concerning the estimation of the vertical velocity. The fluctuations are more attenuated than for the estimation of the streamwise velocity. Indeed, it is important to notice that the scales of images d) and f) are different than the one in image b). For the LSE using 5 sensors, some structures just downstream the airfoil are predicted, but they do not match the original structures. At last, these observations are also true concerning fields from the training set. In overall, the shock location is relatively well estimated. Vortices in the wake are not reconstructed, nor predicted.

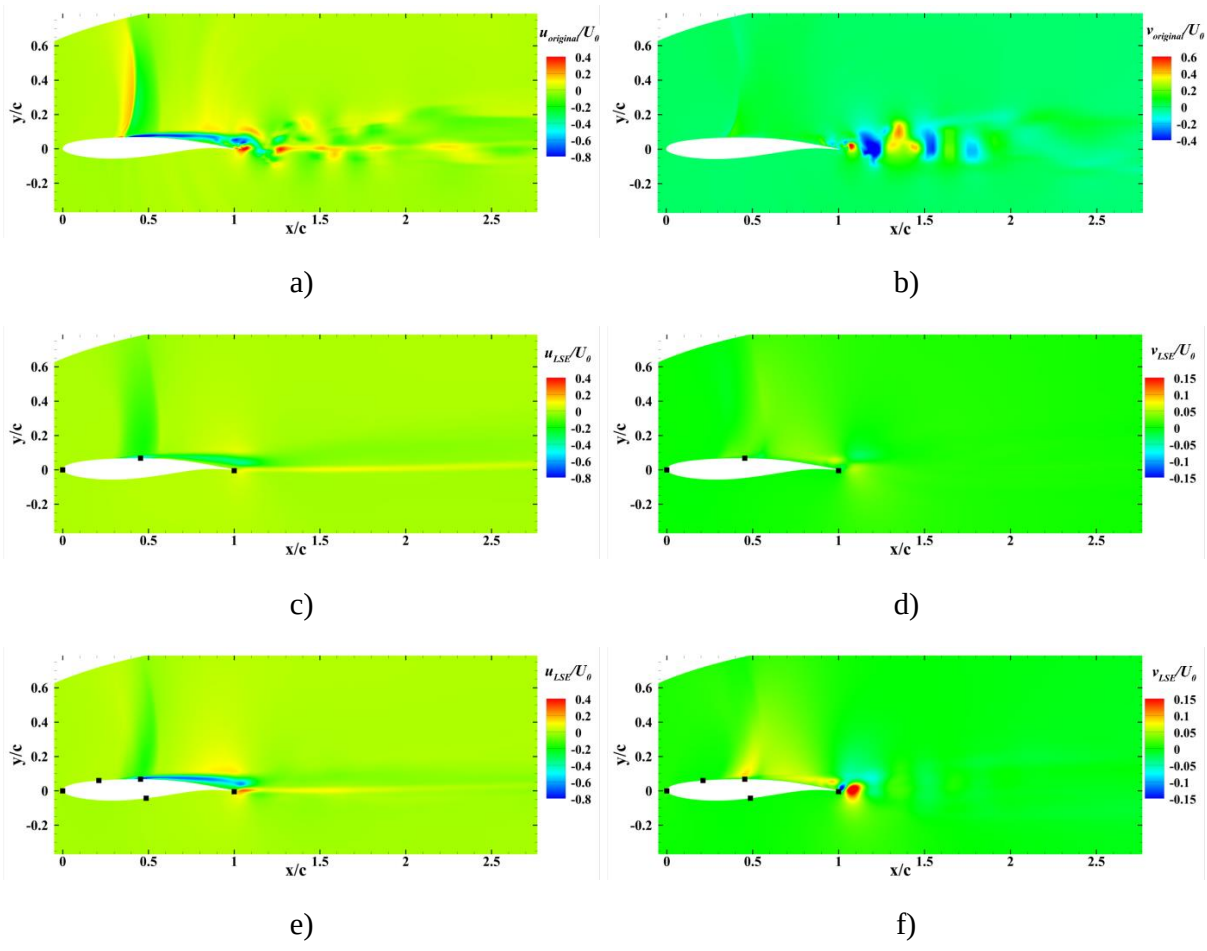


Fig. III.15: Instantaneous fluctuating velocity field normalized by U_0. a), c), e) streamwise component; b), d), f) vertical component. a), b) original field; c), d) LSE prediction with 3 pressure sensors; e), f) LSE prediction with 5 pressure sensors. (Validation set estimation)

In terms of R^2, the results are not as good as for the blunt trailing edge wake. In the 3 sensors case, R^2 is 43.6% and 53.2% is obtained for the "intuitive P 5" set. The normalized mean square error of validation set data estimation is plotted in Fig. III.16 for both velocity components and both 3 and 5 sensor cases. As expected from the observations drawn from the instantaneous field estimation, and from the maps of average correlation module, large errors are found in the wake (for both the streamwise and vertical component). As in the blunt trailing edge case, it is shown that the

maps of average correlation module can identify the areas where the LSE gives poorly accurate estimations. High error levels are also found around the mean shock location. This observation is not in contradiction with the fact that the shock position is well estimated. Indeed, if the shock position is correctly reproduced, velocity levels, upstream and downstream the shock, are not. Differences between the 3 and 5 sensors cases are also important and the addition of the 2 sensors clearly decreases the error levels. However, when comparing the error from both cases, the regions of highest errors appear to be mostly unchanged. It is mostly where the error is about 0.5 in the 3 sensors case, that the error is strongly reduced (below 0.1). Thus the addition of the 2 sensors is not sufficient to recover structures in the wake.

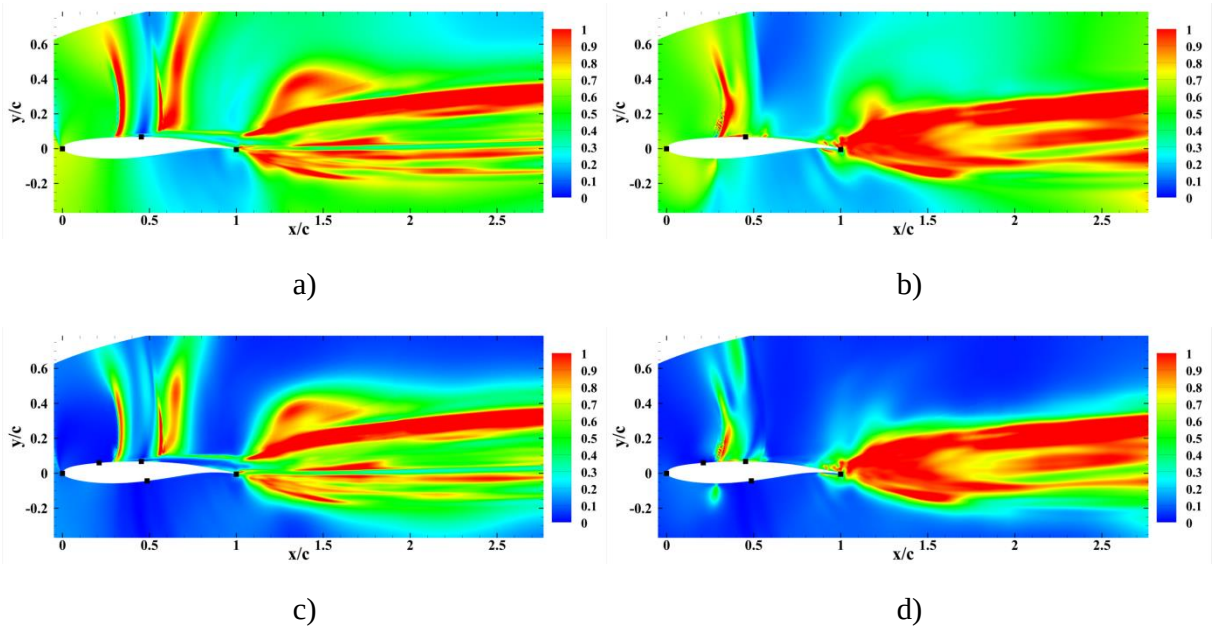


Fig. III.16: Normalized mean square error of the LSE prediction (validation set). a) streamwise velocity component, “intuitive P 3”, b) vertical velocity component, “intuitive P 3”, c) streamwise velocity component, “intuitive P 5”, d) vertical velocity component, “intuitive P 5”.

The use of QSE is now investigated using the same sets of pressure sensors. Once more instantaneous fluctuating velocity fields are displayed in Fig. III.17 and can be compared with those in Fig. II.15. Compared to the LSE prediction, the streamwise velocity field estimated by QSE does not present any strong differences. Streamwise velocity levels are slightly better recovered (it is more visible near the shock). More differences can be observed for the vertical velocity where the estimated levels are less attenuated with QSE than with LSE. At last the differences between the 3 and 5 sensors cases using QSE are not as strong as when LSE is used. The addition of 2 sensors does not improve the prediction as much as in the LSE case.

It is confirmed by values of R^2 which is 58% with 3 sensors and 60.7% with 5. These values show that the QSE estimate is more accurate than the LSE one. The QSE achieves a better accuracy with 3 sensors than the LSE with 5. These observations are consistent with those made in the blunt trailing edge case. The improvement of accuracy can also be observed on maps of the normalized mean square error plotted in Fig. III.18. In this figure, one can see that the error is strongly reduced around the shock and in a smaller amount in the wake.

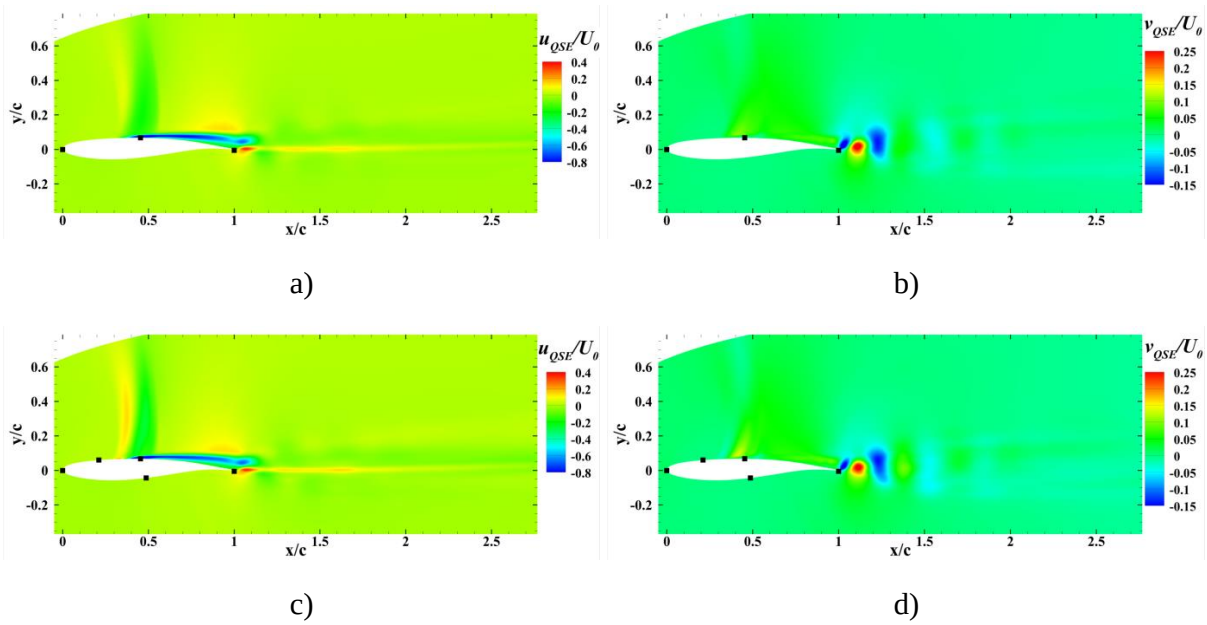


Fig. III.17: Instantaneous fluctuating velocity field normalized by U_0. a), c) streamwise component; b), d) vertical component. a), b) QSE prediction with "intuitive P 3" set; c), d) QSE prediction with "intuitive P 5" set (Validation set estimation)

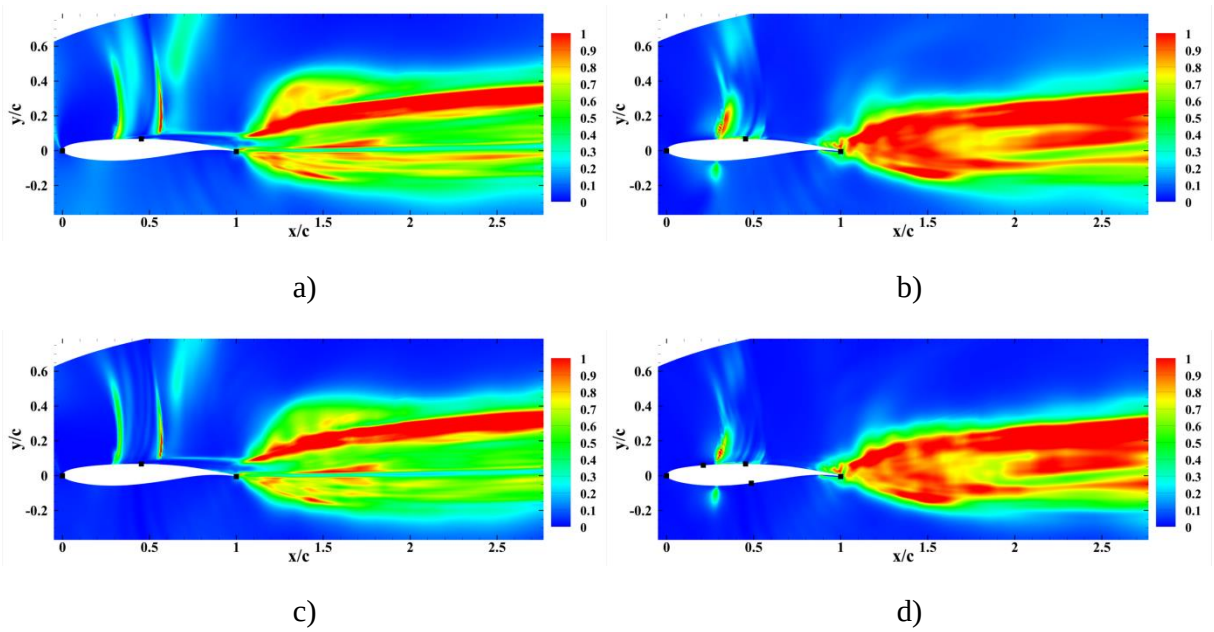


Fig. III.18: Normalized mean square error of the QSE prediction (validation set). a) streamwise velocity component, "intuitive P 3", b) vertical velocity component, "intuitive P 3", c) streamwise velocity component, "intuitive P 5", d) vertical velocity component, "intuitive P 5".

In the literature on LSE and QSE, the quality of the estimation is sometime assessed by looking at the estimation of root-mean-square quantities (see for example [11]). It is important to remark that a good estimation of a root-mean-square quantity does not guarantee the correct estimation of the instantaneous flow. Thus, when one is interested in predicted the temporal evolution of a quantity, evaluating the estimation of its root-mean-square value is clearly not enough.

Vertical root-mean-square velocity fields (original and estimated by LSE, QSE on the validation set) are displayed in Fig. III.19, Fig. III.20 and Fig. III.21. (Streamwise root-mean-square

velocity fields are shown in Annex H.2 Fig. H.2). Vertical root-mean-square velocity is strongly underestimated in the wake (not so much around the shock) whatever the estimation technique and the set of sensors used. However, for the same set of sensors, root-mean-square velocity levels increase when QSE is used instead of LSE. Adding sensors, also leads to an improvement of both reconstruction and prediction of the root-mean-square velocity. These observations are consistent with the general improvement in accuracy observed on the determination coefficient between the several tested situations. The streamwise root-mean-square velocity is, for its part, better recovered than the vertical one (see in Fig. H.2 in Annex H.2).

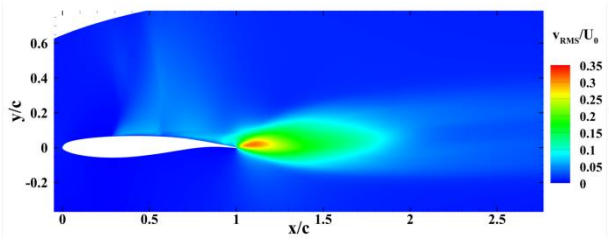


Fig. III.19: Vertical root-mean-square velocity calculated from the validation set

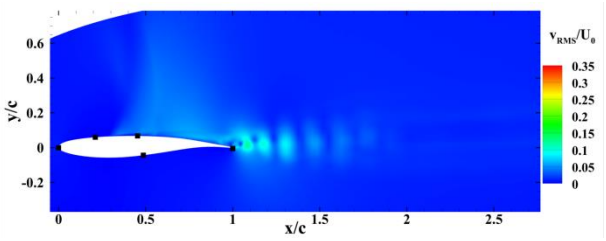


Fig. III.20: Vertical-root-mean square velocity of the LSE estimation of validation set data (5 pressure sensors are used)

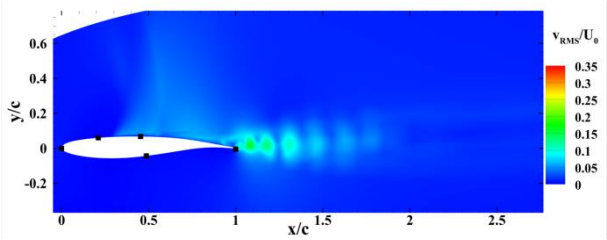


Fig. III.21: Vertical root-mean-square velocity of the QSE estimation of validation set data (5 pressure sensors are used)

2.1.2. Modified Stochastic Estimation (SE-POD)

The POD is computed from the training set snapshots only (so that the POD basis contains no information from the validation set). Thus 4 640 snapshots are used (see Chapter II). Only the streamwise and vertical velocities are considered for the calculation of the POD modes. The energy spectrum of the first hundred POD modes is displayed in Fig. III.22. Compared to the Fig. III.10, the energy content is spread over more POD modes for the OAT15A airfoil flow than for the blunt trailing edge one. The first POD mode accounts for 48% of TKE and the first four for 79%.

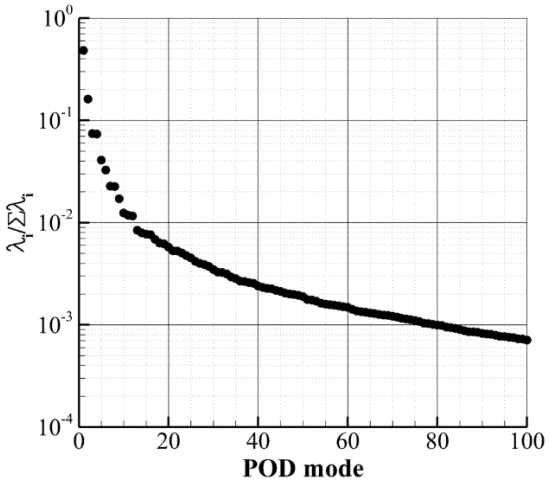


Fig. III.22: Energy spectrum of the first hundred POD modes (OAT15A ZDES)

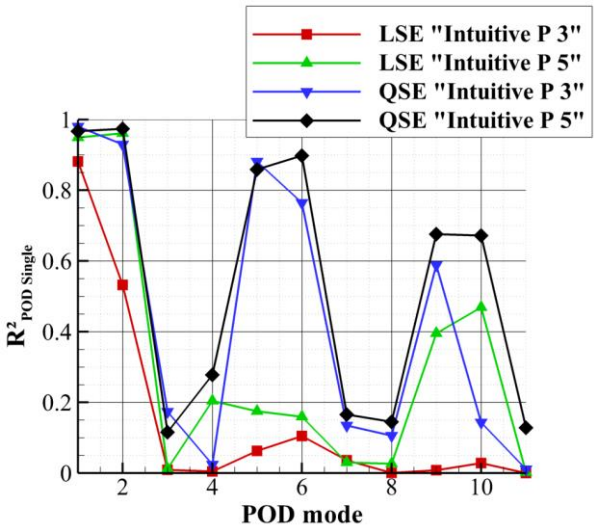


Fig. III.23: Determination coefficient calculated for each POD mode separately on the validation set (OAT15A ZDES)

Four situations, corresponding to different number of POD modes to estimate, are investigated. 2, 4, 25 and 80 POD modes are considered which represents 48%, 79%, 84% and 93% of TKE. Both 3 and 5 sets of pressure sensors are used. Values of R^2 are given in Table III.3 and can be compared with LSE and QSE ones.

As observed in the blunt trailing edge test case, the use of POD does not improve the estimation of the original flow field containing 100% of TKE. It is coherent since the estimation concerns only a limited number of POD modes and thus of TKE. To compute the determination coefficient, it is also possible to take as reference, not the original flow field (containing 100% of TKE), but its projection onto the POD basis that is considered. From the values in Table III.3, one can see that it is always higher than the one obtained from the original flow field. Thus, the estimated flow field is "closer" to the projected field. In addition, this determination coefficient decreases when more POD modes are used. This indicates that POD modes of higher rank are globally estimated with a decreasing accuracy, which is confirmed by the evolution of R^2_{POD} which decreases drastically. In the calculation of R^2_{POD}, each POD coefficient accounts for the same amount in the error since they are normalized. Thus, when several POD coefficients are estimated with a poor accuracy, the error rapidly increases. Actually, R^2_{POD} is almost divided by two when 4 POD modes are considered instead of 2. This seems to indicate that among the POD modes 3 and 4, at least one is not estimated with high fidelity. This is confirmed in Fig. III.23 where the determination coefficient of the first 10 POD modes are plotted. POD modes 1 and 2 are well predicted, especially when 5 sensors are utilized. With the set of 5 pressure sensors $R^2_{POD\ Single} > 90\%$ is achieved. On the contrary, POD modes 3 and 4 are very poorly reconstructed and predicted. With both sets of sensors, $R^2_{POD\ Single}$ is less than 3% for POD mode 3 and less than 20% for POD mode 4. For the "Intuitive P 3" set, all higher POD modes are estimated with $R^2_{POD\ Single}$ less than 15%. For the set of 5 sensors, POD modes 9 and 10 are estimated with $R^2_{POD\ Single}$ higher than 40% but higher order POD modes are also poorly estimated ($R^2_{POD\ Single} < 10\%$).

Sensors set			Intuitive P 3	Intuitive P 5
R^2 from original fields	Prediction	LSE	0.436	0.532
		LSE-POD (2 modes)	0.415	0.495
		LSE-POD (4 modes)	0.416	0.508
		LSE-POD (25 modes)	0.422	0.529
		LSE-POD (80 modes)	0.422	0.529
		QSE	0.58	0.607
		QSE-POD (2 modes)	0.503	0.506
		QSE-POD (4 modes)	0.515	0.527
		QSE-POD (25 modes)	0.579	0.605
		QSE-POD (80 modes)	0.58	0.607
R^2 from POD projected fields	Prediction	LSE-POD (2 modes)	0.799	0.952
		LSE-POD (4 modes)	0.649	0.793
		LSE-POD (25 modes)	0.513	0.644
		LSE-POD (80 modes)	0.467	0.585
		QSE-POD (2 modes)	0.968	0.991
		QSE-POD (4 modes)	0.804	0.823
		QSE-POD (25 modes)	0.704	0.736
		QSE-POD (80 modes)	0.641	0.671
R^2_{POD}	Prediction	LSE-POD (2 modes)	0.713	0.955
		LSE-POD (4 modes)	0.354	0.525
		LSE-POD (25 modes)	0.079	0.158
		LSE-POD (80 modes)	0.025	0.046
		QSE-POD (2 modes)	0.956	0.969
		QSE-POD (4 modes)	0.52	0.578
		QSE-POD (25 modes)	0.24	0.316
		QSE-POD (80 modes)	0.076	0.104
$\langle VIF \rangle$		LSE	94.8	1.40×10^{5}
		QSE	3.65×10^{7}	3.78×10^{10}

Table III.3: Determination coefficients comparison between several SE (OAT15A ZDES) (calculated for the validation set)

When QSE-POD is used, the same conclusions can be drawn concerning the comparison between QSE and QSE-POD. In addition, QSE-POD leads to higher determination coefficients than LSE-POD in all situations. More interesting are the observations made on the estimation of POD mode separately when using QSE instead of LSE (see Fig. III.23). If the QSE improves, in overall the accuracy of the reconstruction and prediction, the improvement is very different from one POD mode to another. For instance, $R^2_{\text{POD Single}}$ goes from less than 20% for POD modes 5 and 6 (whatever the set of sensors) to more than 75% (for both set of sensors) when using QSE instead of LSE. Determination coefficient also increases strongly for POD mode 9 when QSE is used. Taking into account the quadratic term in the Taylor expansion is therefore very important to estimate some POD modes. In comparison, increases of $R^2_{\text{POD Single}}$ for POD modes 3, 4, 7 and 8 are small and they are all estimated with a determination coefficient lower than 30%. Therefore, for some POD modes, the addition of quadratic information from the sensors is not enough to get an accurate prediction.

Some POD modes are clearly more difficult to estimate with a limited number of pressure sensors using SE methods. This will be better analyzed in the next chapter (see Chapter IV.2).

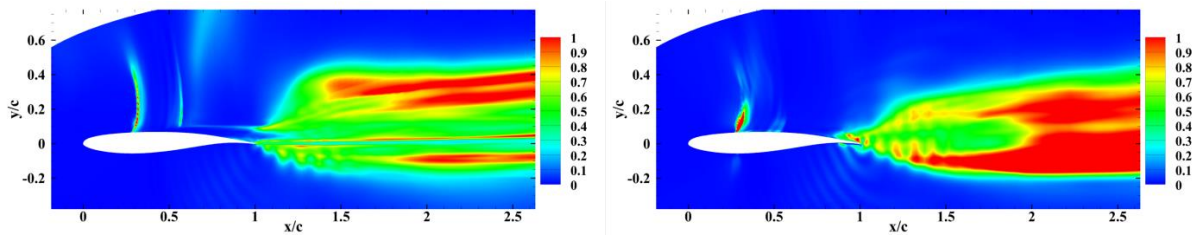
2.2. Multi-time-delay Stochastic Estimation (MTD-SE)

As explained in the literature review (see Chapter I.2), recent developments on MTD-SE show the existence of an optimal time window for the choice of delays. Using delayed information outside this window (more in the past) only improved the reconstruction but not the prediction accuracy. Knowing this optimal window, one has still to choose the time step (or equivalently the number of delays). In the literature, to find these parameters several situations are tested and the best is kept. In general, the optimal window is of the order of magnitude of the main phenomena period (for instance, in the case of vortex shedding, it is of order of the vortex period). Then, if a constant time step is considered between the delays (chosen in the optimal time window), the accuracy of the prediction is mostly improved by shortening the time step (and thus increasing the number of delays taken in the time window).

Using both "intuitive P 3" and "intuitive P 5" set of sensors, MTD-LSE is investigated. Several constant time steps between delays are considered, as well as different numbers of delays. Fig. III.25 shows the evolution of R^2 (from the validation set) with the number of delays for several constant time steps between delays. The values are obtained using Tikhonov regularization. Without applying Tikhonov regularization, validation set R^2 mostly decreases when more delays are used and the prediction accuracy is never as good as when Tikhonov regularization is employed. On the contrary, without regularization, the reconstruction accuracy improves when more delays are used.

In Fig. III.25, R^2 increases when the number of delays increases and reaches a plateau when the maximal time delay approaches the shock oscillation period (T_{osc}). In some cases, R^2 reaches a maximum when the maximal delay is between 0.6 – 0.9 T_{osc}. The optimal time window goes from 0 to 0.5 – 1 T_{osc}, which is consistent with observations from the literature (see Tu et al. [65], Clark et al. [23], Hosseini et al. [66]).

Both reconstruction and prediction of the velocity field quality can be improved using MTD-LSE compared to LSE and QSE. R^2 obtained with the best MTD-LSE on the validation set are given in Table III.4. On the training set, R^2 up to 78% are obtained. On the validation set, R^2 reaches 70% using the "intuitive P 5" set and 70 delays every 8.62×10^{-3} T_{osc}. In comparison, R^2 of the LSE prediction, with the same set of sensors, is only 53% and R^2 of the QSE 61%. The normalized mean squared error of the MTD-LSE using "intuitive P 5" set and 70 delays every 8.62×10^{-3} T_{osc} is shown in Fig. III.24. The error is decreased mainly in the wake close to the airfoil compared to LSE and QSE.



a) b)

Fig. III.24: Normalized mean square error of the MTD-LSE prediction using "intuitive P 5" (validation set). a) streamwise velocity component, b) vertical velocity component

	“Intuitive P 3”	“Intuitive P 5”
R^2 (%)	69	70.4

Table III.4: Determination coefficients of the best MTD-LSE prediction using the two sets of pressure sensors

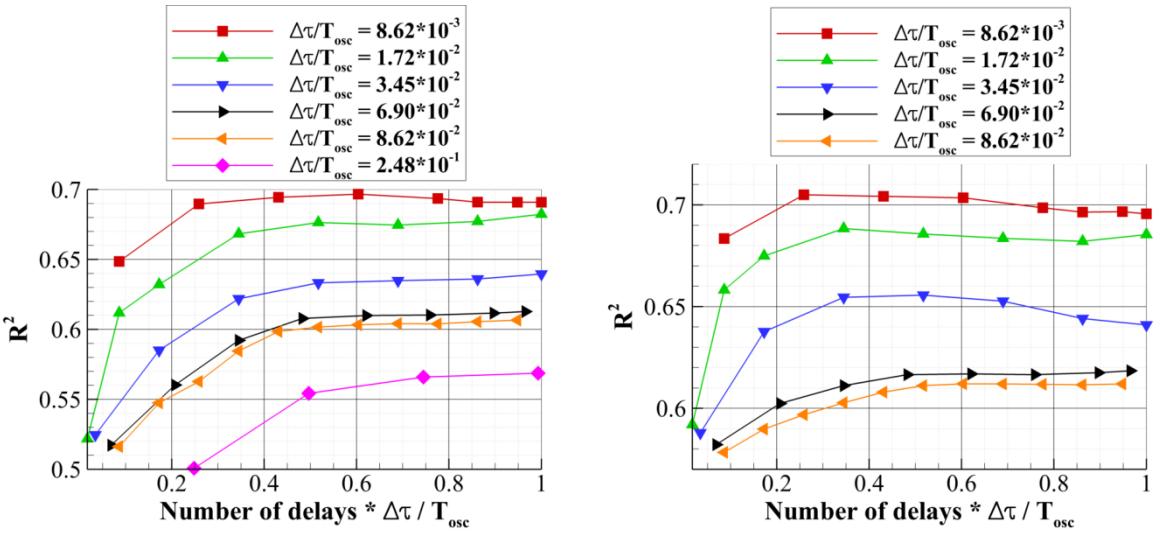


Fig. III.25: Validation set determination coefficients as function of the number of delays for MTD-LSE using 3 (a) and 5 (b) pressure sensors

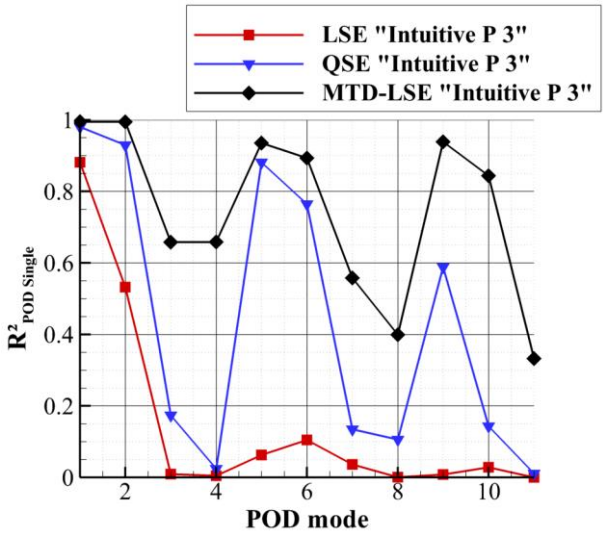


Fig. III.26: Determination coefficient calculated for each POD mode separately on the validation set (blunt trailing edge). The MTD-LSE uses 70 delays every 8.62×10^{-3} T_{osc} in the past and Tikhonov regularization.

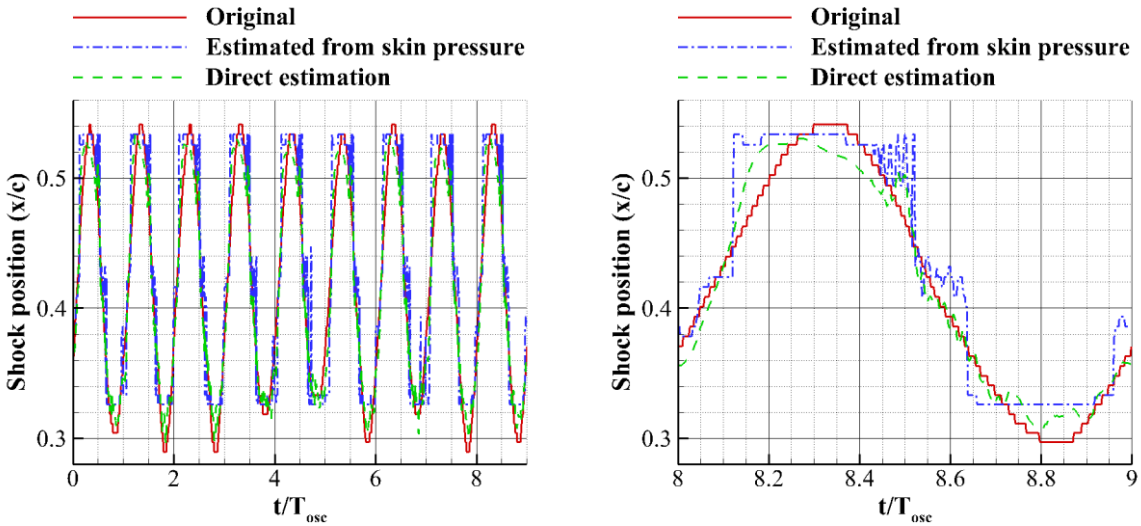
To better explore the impact of MTD methods, $R^2_{\text{POD Single}}$ of the 10 first POD modes for the best MTD-LSE situation using the set of 3 pressure sensors is plotted in Fig. III.26. LSE and QSE results are also shown for comparison. MTD-LSE leads for all of the 10 first POD modes to a better prediction accuracy than LSE or QSE. Especially, the prediction of mode POD 3, 4, 7 and 8 is greatly improved, with $R^2_{\text{POD Single}}$ higher than 40% whereas it was lower than 30% with LSE and QSE (using both set of sensors). Clearly, some information, that could not be predicted using only

single-time, is predicted using several delays in the past. This will be addressed more precisely in the next chapter (Chapter IV).

2.3. Estimation of the shock position

The possibility to predict the shock position (on the airfoil wall) is quickly investigated. Indeed, it is possible to think of applications for which the knowledge of the entire velocity field is not necessary and more limited information, like the shock position, might be sufficient. That is the case when applying control to the buffet phenomenon. In what follows the shock position on the airfoil wall is defined using the pressure information at the wall. The shock position is the position at which the absolute value of the spatial pressure derivative, on the suction side, is maximum. To compute the spatial pressure derivative, a second order centered scheme is used and only the x coordinate is considered. The shock position evolution in time is displayed in Fig. III.27 (original curve). The steps that the curve presents are due to the spatial discretization of the airfoil skin.

Two methods are studied to predict the shock position. The first one is to predict the pressure at the skin of the airfoil and then calculate the shock position. The second one is to predict directly the shock position. Only the set of 5 pressure sensors ("intuitive P 5" is considered). The results are displayed in Fig. III.27 and show that the shock position is qualitatively well estimated by the LSE in both cases. Directly estimating the shock position using LSE seems to lead to slightly better results and that is confirmed quantitatively by looking at the determination coefficient values. R^2 is 99.8% when direct estimation is used and 90.3% when the estimation of the skin pressure is used. Therefore, it seems that if one wants to estimate a variable extracted from the field, it is better to directly estimate it, than to estimate the field and then recalculate the variable. At last, if the field is predicted with an average accuracy, the shock position can still be predicted by LSE, using the same conditional event, with a very high accuracy.



a) b)

Fig. III.27: Shock position as function of time. Comparison between the original data and the estimations by LSE through skin pressure estimation or by direct estimation. (b) zoom on the validation period.

2.4. Summary

Comparing LSE, QSE and MTD-LSE leads to the same conclusions as in the blunt trailing edge case. MTD-LSE (with Tikhonov regularization) outperforms the LSE and QSE. QSE also improves the reconstruction and prediction accuracy compared to LSE. The use of LSE-POD and QSE-POD does not improve the estimation of the entire flow field. The addition of conditional events is also shown to improve the estimation accuracy.

In overall, the estimation is not as good as for the blunt trailing edge and the highest R^2 obtained with MTD-LSE is only 70%. Looking at the estimation of POD modes, it appears that some are estimated with a much higher accuracy than others. The study of the phenomenon is conducted in Chapter IV.

Also the possibility to estimate the shock position was investigated. The study showed that it is better to directly estimate it than to estimate the field (or skin pressure in this case) and then calculate it. It also showed that a very high accuracy was obtained while the prediction accuracy of the velocity field was average.

3. Backward Facing Step flow case

In this section, the experimental data from the low speed PIV campaign are first used to test the SE. Then SE is also applied to the numerical database. Only the 30 $m.s^{-1}$ case is detailed here. Results obtained for 20 and 25 $m.s^{-1}$ of upstream velocity are however summarized in Table III.7, Table III.8, Table III.9 and Table III.10. Some figures are also given for illustrations in annex H.3.

3.1. Application to the experimental database

3.1.1. Single-time Stochastic Estimation

In this part, SE is tested for the reconstruction and prediction of the BFS low speed PIV database. The 17 pressure sensors, synchronized to the snapshots, are used.

3.1.1.1. Linear Stochastic Estimation (LSE) and Quadratic Stochastic Estimation (QSE)

As previously, a first evaluation step can be to look at maps of the average correlation module between the sensors and the field to estimate (see Fig. III.28). Compared to Fig. III.14 levels of correlations are very small in the BFS case. A region of higher correlations is nevertheless observed downstream the step, above the recirculation region, at about the reattachment point. Lower errors are therefore expected in this region, but one can expect the LSE to perform poorly in overall.

An example of an instantaneous field (from the validation set) and its estimation is shown in Fig. III.29 for the streamwise velocity and in Fig. III.30 for the vertical velocity. One thing to notice is that in order to see the predicted structures, two different scales are used, one for the original field and one for the estimated one. From this fact it is obvious that predicted fluctuations are greatly lessened. They are almost 10 times lower than the original one. The original field is composed of numerous small scale structures of high levels of fluctuations. On the contrary, in the LSE prediction, only large structures of lower levels of fluctuations appear. In these figures, it is very difficult to see a relation between the original and estimated fields. The accuracy of the LSE is very low, which is confirmed by the determination coefficient: 8.4% on the training set and 7.9% on the validation set. The normalized mean square error (calculated on the validation set) is shown in Fig. III.31. The error is extremely high (higher than 0.9) in most of the field. As hinted by the study of

the average correlation module, lower errors are found above the recirculation region, at the reattachment point. However, the lowest error there is still about 0.5 which remains high.

The prediction by QSE of the same instantaneous flow field as in Fig. III.29 and Fig. III.30 is plotted in Fig. III.32. As one can observe, the scales of the predicted fields are still different between the original and the QSE prediction. But the extrema of the QSE scales are higher than those of the LSE scales. The predicted structures by the QSE are recovered with higher fluctuation levels. However, the comparison between the LSE and QSE predicted flow field shows that the predicted structure shapes and positions are almost unchanged. Both predictions are in fact very similar. The normalized mean square error of the QSE prediction is plotted in Fig. III.33. The figure is completely similar to Fig. III.31. No real improvement can be observed by using the QSE. The QSE determination coefficient is 11.6% on the training set, but only 6.9% on the validation. This difference of R^2 between the training set and validation set, which was not observed in the blunt trailing edge case and OAT15A case, is symptomatic of an overfitted model. Going back to the equations of the Chapter II.1.1, it was shown that using single-time QSE or LSE leads to the resolution of a matrix system of the same form (see equation (II.8)) for the calculation of the coefficients. From a simple mathematical point of view, it is possible to consider the QSE as being an LSE using more conditional events. These additional events are in fact the $\frac{N\times(N-1)}{2}$ permutations $\langle \mathbf{E}(y_i)\mathbf{E}(y_j)\rangle$ (with order). With 17 sensors, the QSE holds in total 153 coefficients. Therefore, the QSE model is greatly larger and complex than the LSE one which makes it more susceptible to suffer from overfitting. Considering each permutation $\langle \mathbf{E}(y_i)\mathbf{E}(y_j)\rangle$ as an independent conditional event, it is possible to calculate a different VIF for the QSE. The QSE VIF is 4.7 instead of 2 in LSE. The there are more collinearities in the conditional event matrix. In addition, it was observed in Chapter II.3.2.4.1 that the statistical moments used to compute the QSE coefficients were not well converged on the training data set. This is another factor which explains that QSE suffers from overfitting in this test case. It is probably the main factor since the difference in VIF between LSE and QSE is not so important. The use of the Tikhonov regularization, to overcome this overfitting, shows only a very slight improvement (less than 1% on the validation set determination coefficient).

The estimation of the Root-Mean-Square (RMS) velocity is now evaluated. Velocity RMS fields are displayed in Fig. III.34 and Fig. III.35. In these figures, only validation set data are used. The original field can be compared to the estimated field obtained by LSE and QSE with 17 pressure sensors. In overall the predicted RMS are lower than the original ones for both velocity components. It is important to notice that scales are not the same for the original RMS and the predicted ones. Levels are approximately divided by 2. The RMS velocity is however better predicted by QSE than LSE. This has been observed in several works on SE (see for example Murray et al. [11], Aranajatesan et al. [40]). It is interesting since it was previously shown that QSE suffers from overfitting and the prediction accuracy of QSE was here lower than the accuracy of LSE (in term of determination coefficient). In short, while the quality of the prediction of instantaneous velocity field is deteriorated when using QSE, the RMS velocity prediction quality is improved. The explanation is that the levels of the predicted fluctuation velocity are higher in QSE than in LSE. Since the prediction is inaccurate, the increase in levels of velocity fluctuations results in a higher error between original instantaneous field and its prediction (in general). The prediction accuracy is deteriorated. However, the RMS velocity is linked to the levels of fluctuations of the velocity. Thus when they increase, the RMS velocity increases. Therefore, the RMS velocity is less underestimated by QSE than LSE. An interpretation is that, in our case, the use of QSE leads to the prediction of higher fluctuations of velocity than LSE, globally in the correct area, but not particularly at the correct time. In these conditions, comparing LSE and QSE, and evaluating their accuracy through the analysis of estimated RMS quantity is clearly not correct.

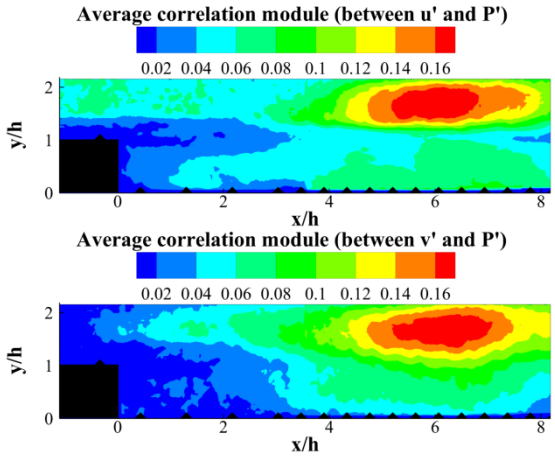


Fig. III.28: Average correlation module between the streamwise velocity (top) or the vertical velocity (bottom) and the set of 17 pressure sensors

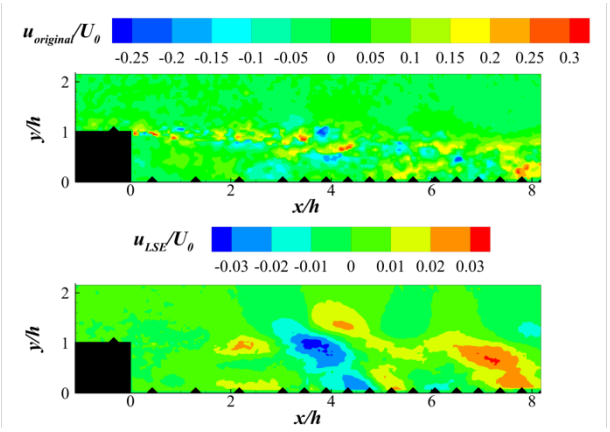


Fig. III.29: Instantaneous fluctuating streamwise velocity field. Top: original, bottom: LSE prediction.

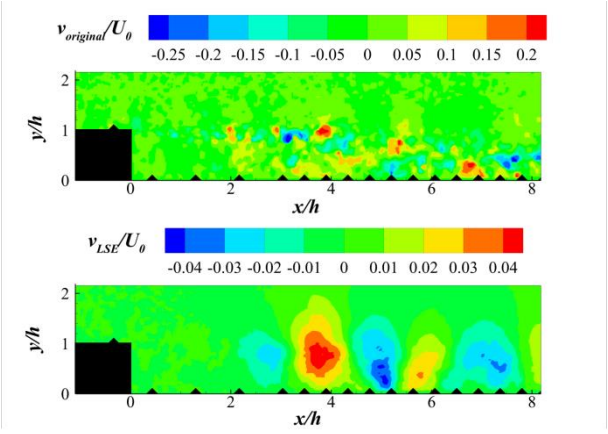


Fig. III.30: Instantaneous fluctuating vertical velocity field. Top: original, bottom: LSE prediction.

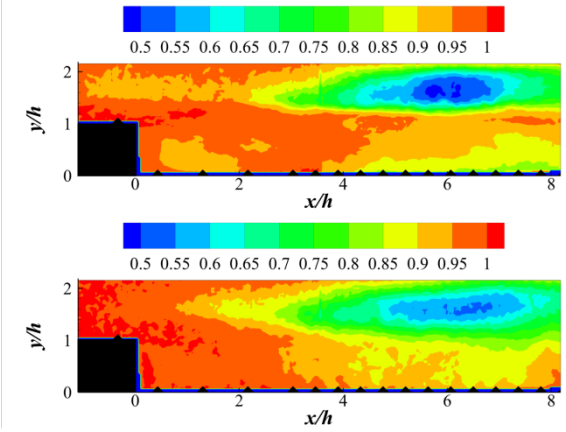


Fig. III.31: Normalized mean square error of the LSE prediction for the streamwise velocity (top) and the vertical velocity (bottom)

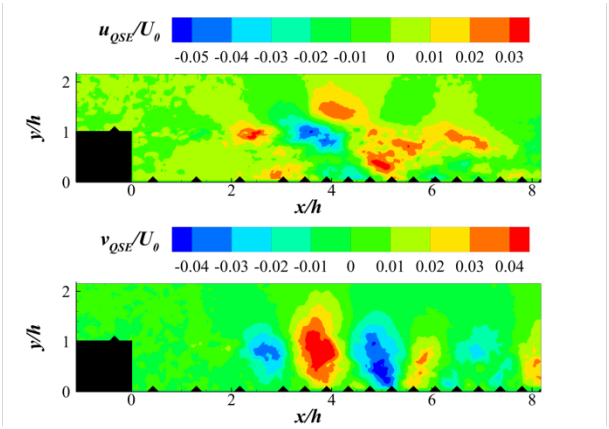


Fig. III.32: Instantaneous fluctuating velocity field predicted by QSE. Top: streamwise component, bottom: vertical component

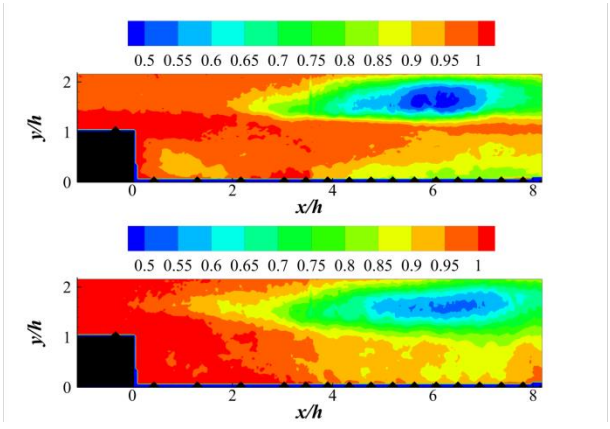


Fig. III.33: Normalized mean square error of the QSE prediction for the streamwise velocity (top) and the vertical velocity (bottom)

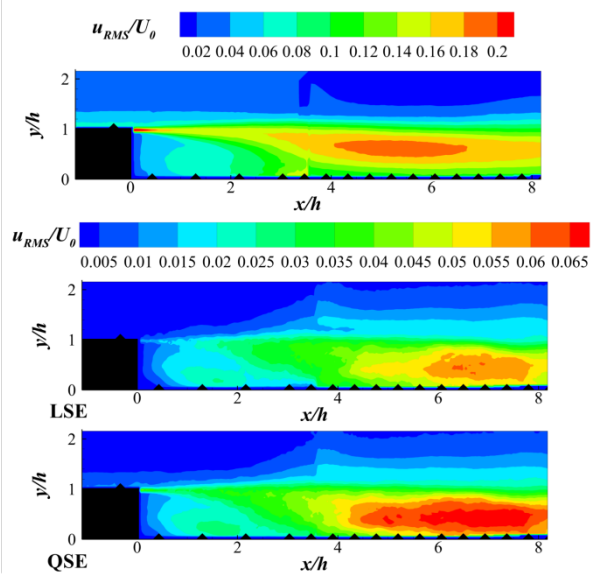


Fig. III.34: Streamwise root-mean-square velocity; top: original; middle: LSE estimation; bottom: QSE estimation (validation set)

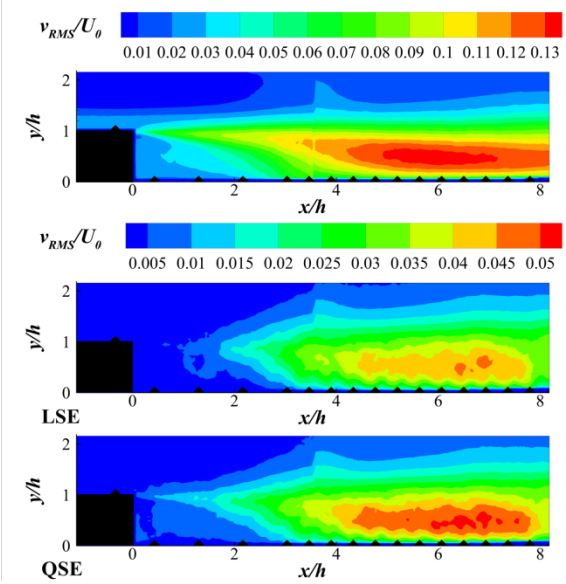


Fig. III.35: Vertical root-mean-square velocity, top: original; middle: LSE estimation; bottom: QSE estimation (validation set)

3.1.1.2. Modified Stochastic Estimation (SE-POD)

Once more the POD is computed from the training data set only. The energy spectrum of the POD is plotted in Fig. III.36. Compared to the spectrum obtained for the OAT15A or the blunt trailing edge case, the TKE is clearly spread over a larger number of POD modes. The first POD mode accounts only for about 11% of the total TKE. The first 20 POD modes are necessary to retain about 50% of the TKE and 652 POD modes to retain 90% of the TKE. In this test case, the flow dynamics cannot be captured and modeled with a small number of POD modes.

The prediction of the same instantaneous fluctuating velocity field as in Fig. III.29, Fig. III.30 and Fig. III.32, using LSE-POD is shown in Fig. III.38. 92 POD modes are used, which retain about 70% of the TKE. The original velocity field projected on the POD basis is also displayed. Compared to the LSE prediction, the predicted structures remain unchanged in shape and location. The main difference is that the levels of predicted fluctuating velocity are decreased in the LSE-POD case. It seems logical since less POD modes are considered (using all the POD modes is identical to executing a direct LSE). Indeed, the projection of the original field on the POD basis has lower levels of velocity fluctuations. Also, the visualization of the POD projection of the original velocity field visualization shows more similitudes with the predicted field. For instance, sign change of the streamwise fluctuating velocity at about $x/h = 4$ and $y/h \in [\![1,2]\!]$ is predicted. The succession of positive and negative u near the wall, between $x/h = 5$ and $x/h = 8$, seems also qualitatively recovered. Therefore, the predicted velocity field holds some qualitative information on the original field. But, when this predicted information is evaluated using previously defined error criteria, it appears particularly poor.

Determination coefficients of several LSE-POD estimations can be compared in Table III.5. For most situations, the use of LSE-POD instead of LSE does not increase R^2 computed with the original field (100% TKE) as reference. This was observed in the two previous test cases. However, in the case where 652 POD modes are used, a slight increase of R^2 is observed. In addition, the increase is higher for the validation than for the training set. This evolution seems to indicate that the reconstruction and the prediction of POD modes of rank higher than 652 actually penalized the estimation. The estimation by LSE of those POD modes is so inaccurate that it deteriorates the

overall quality of the entire velocity field estimation. In this case, using the POD owns a new advantage. It allows leaving out POD modes which estimation deteriorates the estimation of the entire velocity field. Nevertheless, in the present case, putting aside the highest POD modes does not lead to a significant accuracy improvement (less than 0.1% on R^2). Concerning R^2 calculated with the POD projection of the original field as reference and of R^2 directly calculated from the POD coefficients, their evolution is in perfect agreement with observations made for the OAT15A and the blunt trailing edge case. As more POD modes are used, they decrease which means that POD modes of higher ranks are globally estimated with lower accuracy.

$R^2_{\text{POD Single}}$ of the prediction of the first 3000 POD modes by LSE is plotted in Fig. III.37. For all POD modes, $R^2_{\text{POD Single}}$ is less than 20% which is consistent with the low value of R^2 obtained from the comparison with the original fields. In addition, $R^2_{\text{POD Single}}$ decreases rapidly, going from 20% to less than 10% for POD modes of rank higher than 11, and less than 5% for POD modes of rank higher than 44. At last, from POD modes 500 approximately, most of $R^2_{\text{POD Single}}$ values are negative. It is consistent with previous observation showing that using more than 652 POD modes deteriorates the prediction quality.

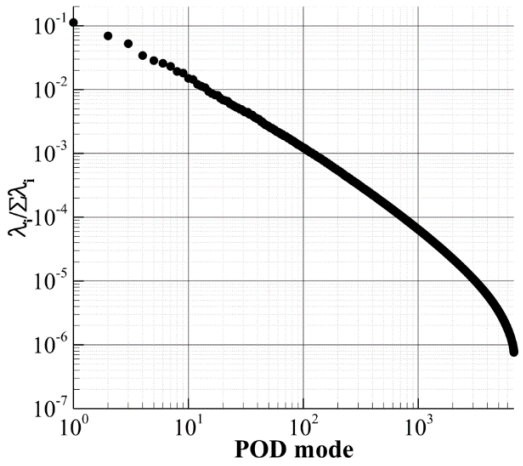


Fig. III.36: Energy spectrum of all POD modes (BFS: low speed PIV database)

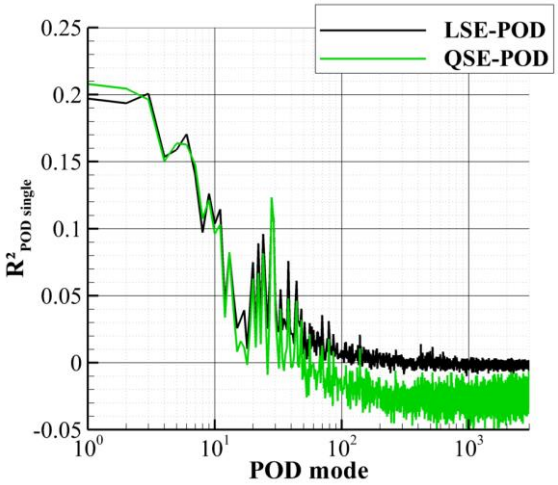
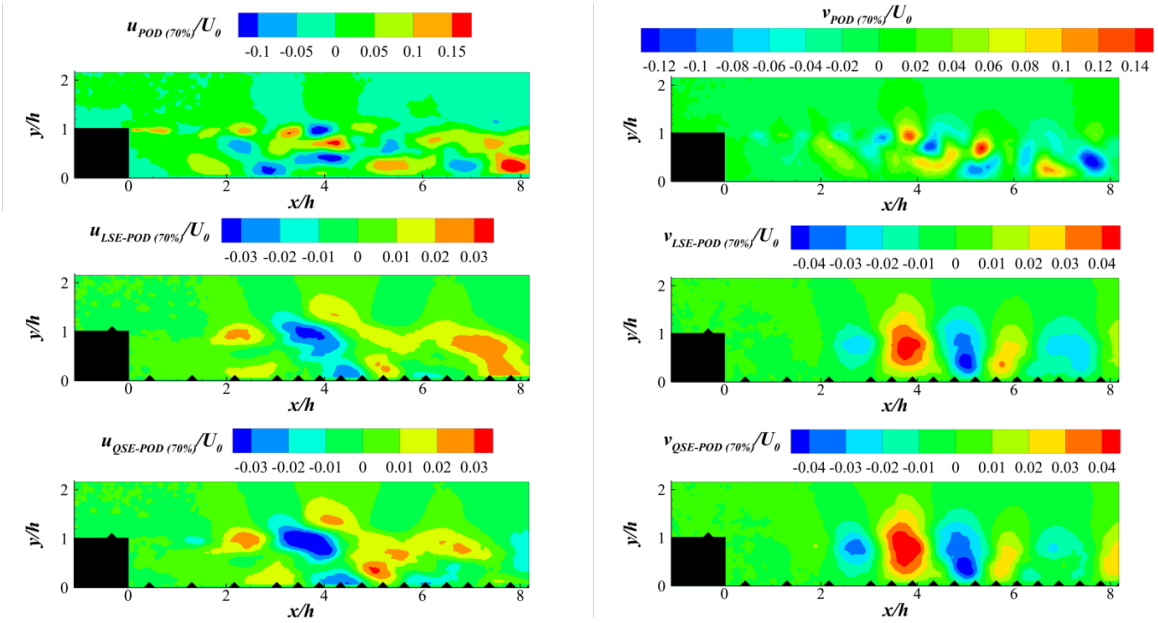


Fig. III.37: Determination coefficient calculated for each POD mode separately on the validation set using LSE and QSE (BFS: low speed PIV)

QSE-POD is also studied. An instantaneous fluctuating velocity field is shown in Fig. III.38. Compared to the LSE-POD predictions, fluctuating velocity levels are slightly higher in some areas. Globally, the structures are very similar. As explained previously, the increase of predicted fluctuating levels of velocity results in higher error and thus lower R^2. Values of R^2 are given in Table III.6 and can be compared with those of LSE in Table III.5. All values calculated on the training set are higher when using QSE. Thus the QSE improves the reconstruction accuracy. The improvement is however not so significant (less than 3%). For validation set values, they are lower for QSE than for LSE (except in the case where the POD basis used retains only 32% of TKE). Thus the overfitting is also observed when QSE is associated with POD. However, the used of POD seems to be able to compensate a bit for the overfitting problem. First, when only 6 POD modes are considered (32% of TKE), no overfitting is observed. Then, when the number of POD modes considered is increased, R^2 goes by a maximum. Such behavior seems also to induce that the first POD modes prediction does not suffer from overfitting (or not as much as higher POD modes). This is confirmed when looking at Fig. III.37. The prediction of the first 2 POD modes is more accurate with QSE than LSE. But from the 10$^{\text{th}}$ POD modes, most of POD modes are better predicted using LSE than QSE. The overfitting is particularly highlighted in the figure from POD modes 50 approximately,

with R^2 decreasing down to negative values. Therefore, using about 50 POD modes for the QSE-POD should lead to the highest R^2 value on the validation set. It is the case, with R^2 = 7.65%. This value remains, however, lower than the LSE one.



a) b)

Fig. III.38: Instantaneous fluctuating a) streamwise, b) vertical velocity field. Top: projection of the original field onto the first 92 POD modes, middle: LSE-POD prediction, bottom: QSE-POD prediction.

		LSE	LSE-POD (90%, 652 modes)	LSE-POD (70%, 92 modes)	LSE-POD (50%, 20 modes)	LSE-POD (32%, 6 modes)
R^2 from original fields	Training	8.44	8.45	8.33	7.27	6.03
	Validation	7.91	7.96	7.91	7.61	5.86
R^2 from POD projected fields	Training	-	9.38	11.9	15.1	18.7
	Validation	-	9.24	11.7	14.9	18.7
R^2_{POD}	Training	-	1.02	4.55	10.2	17.8
	Validation	-	0.72	4.28	9.89	17.8

Table III.5: Determination coefficients comparison between LSE and several LSE-POD (BFS: low speed PIV; values are expressed in %)

		QSE	QSE-POD (90%)	QSE-POD (70%)	QSE-POD (50%)	QSE-POD (32%)
R^2 from original fields	Training	11.7	11.4	10.8	9.6	7.4
	Validation	6.91	7.15	7.58	7.31	6.01
R^2 from POD projected fields	Training	-	12.7	15.5	19.1	22.9
	Validation	-	8.31	11.2	15	19.1
R^2_{POD}	Training	-	3.4	7.3	13.6	21.5
	Validation	-	-1.8	2.5	9.3	18.1

Table III.6: Determination coefficients comparison between QSE and several QSE-POD (BFS: low speed PIV; values are expressed in %)

3.1.2. Multi-Time-Delay Stochastic Estimation (MTD-SE)

As for the OAT15A airfoil case, several sets of delays are used and compared to found the one leading to the most accurate prediction. Inside a set of delays, the time step between consecutive delays is constant. Training and validation sets R^2 evolutions with the number of delays for several constant time steps between delays are plotted in Fig. III.39. Training set values are obtained without Tikhonov regularization, while validation set ones are obtained using regularization. As previously, if no regularization is used, lower R^2 are reached for the validation set. As observed in the literature, the accuracy of the reconstruction improves when more and more delays are used. In addition, the reconstruction accuracy improves when the constant time step between delays decreases. R^2 reaches more than 45% which is far higher than the 11.7% obtained with QSE.

On the opposite, the prediction accuracy goes by a maximum, when the number of delays increases, which defines an optimal time window for the choice of delays. This was already discussed in Chapter III.2.2. In the tested cases, the optimal time window ranges from 0 to 0.6 – 0.8 $T_{shedding}$ (see Fig. III.39), which is consistent with values found in the OAT15A airfoil test case and in the literature. In overall, higher values of R^2 are reached when the constant time step between the delays decreases. However, between the three tested situations, $\Delta\tau = 1.10^{-2}\ T_{shedding}$ and $\Delta\tau = 5.10^{-2}\ T_{shedding}$ leads to very similar results, with slightly higher determination coefficient in the latter case. Therefore, it seems that at some point reducing the constant time step between the delays does not improve the quality of the prediction. That may be due to the fact that data at consecutive delay become too correlated. The conditioning of the system (II.7) may increase and counterbalance the improvement that the additional information gives. At last, the highest determination coefficient obtained with MTD-LSE, on the validation set, is only 9.3% which is less than 2% better than the LSE. The overall prediction accuracy of the MTD-LSE is still very poor in this test case. The prediction of the same instantaneous velocity field as in Fig. III.29 and Fig. III.30, using MTD-LSE with 75 delays taken every $1\times10^{-2}\ T_{shedding}$, is shown in Fig. III.40. Compared to the LSE and QSE prediction, the MTD-LSE presents some differences in structures for the streamwise velocity, but not so much for the vertical velocity. Some similitudes are observed, mainly the succession of positive and negative streamwise velocity, but the shape of the structures is changed and the structures are more fragmented. Levels are not particularly higher in the MTD-LSE than in the LSE case. Qualitatively, it is difficult to judge if the predicted streamwise velocity field in Fig. III.40 is closer to the original field than the LSE or QSE prediction. The normalized mean square error is displayed in Fig. III.41 and can be compared with the one calculated from LSE and QSE results in Fig. III.31 and Fig. III.33. It is easily seen that the error decreased in the area where it was already the lowest in LSE and QSE (above the recirculation at x/h about 4 to 8). But for the region upstream $x/h \approx 4$, the error seems mostly increased. The impact of the MTD-LSE is not uniform. In some regions, the prediction accuracy is improved, in some others it is deteriorated.

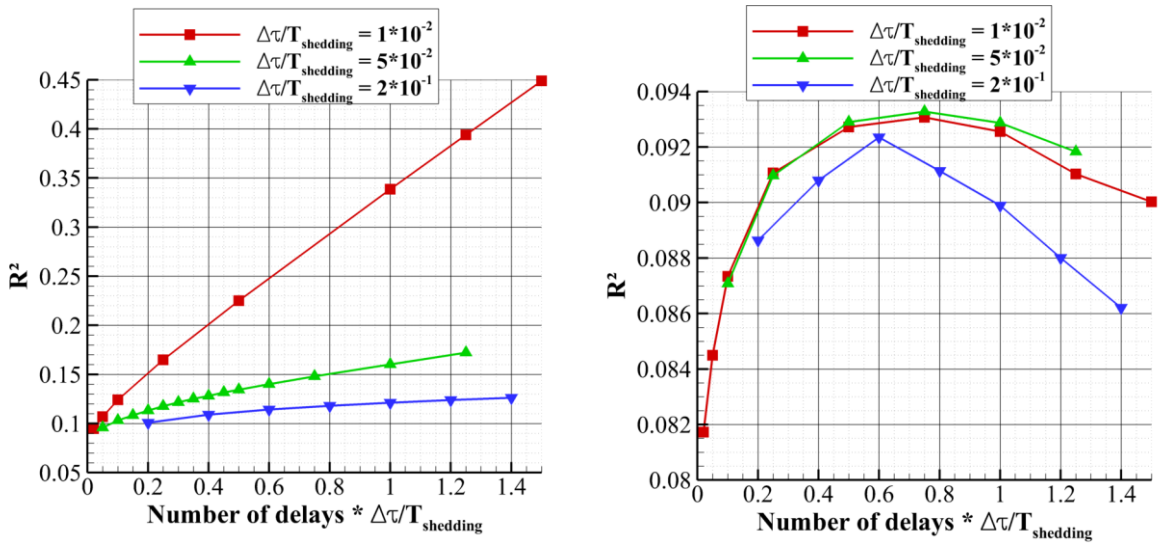


Fig. III.39: Determination coefficients as function of the number of delays for MTD-LSE reconstruction of the training set (a) and prediction of the validation set (b)

At last, the impact of using MTD-LSE on the reconstruction and prediction of POD modes is investigated. The determination coefficients of the first 90 POD modes predicted by MTD-LSE-POD (using 75 delays every 1×10^{-2} $T_{shedding}$) are plotted in Fig. III.42 and can be compared with the LSE and QSE prediction. For some POD modes, $R^2_{\text{POD Single}}$ is increased by more than 5%. But for other POD modes, the improvement is marginal. For instance, $R^2_{\text{POD Single}}$ increases by less than 1% for the POD mode 23 and by about 5% for the POD mode 24. In addition, for POD modes of rank higher than 50, $R^2_{\text{POD Single}}$ is, at best, increased marginally. Thus, the MTD-LSE in the BFS case seems to only improve the prediction of some POD modes and the improvement is not so significant. MTD-LSE is not enough to achieve high prediction accuracy as in the OAT15A airfoil or blunt trailing edge cases.

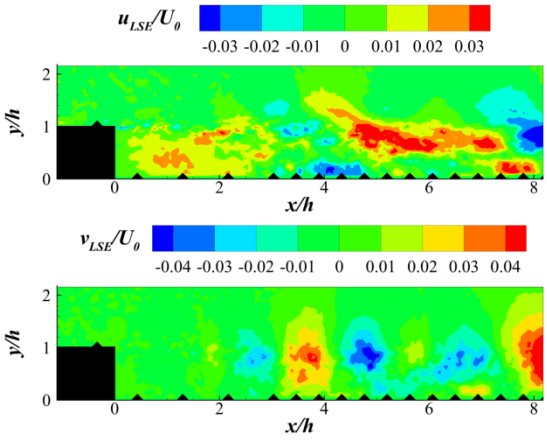


Fig. III.40: Instantaneous fluctuating streamwise (top) and vertical (bottom) velocity field predicted by MTD-LSE

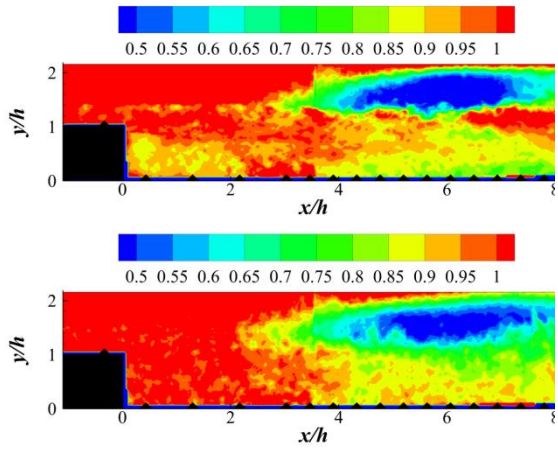


Fig. III.41: Normalized mean square error of the MTD-LSE prediction for the streamwise velocity (top) and the vertical velocity (bottom) (75 delays every 1×10^{-2} $T_{shedding}$)

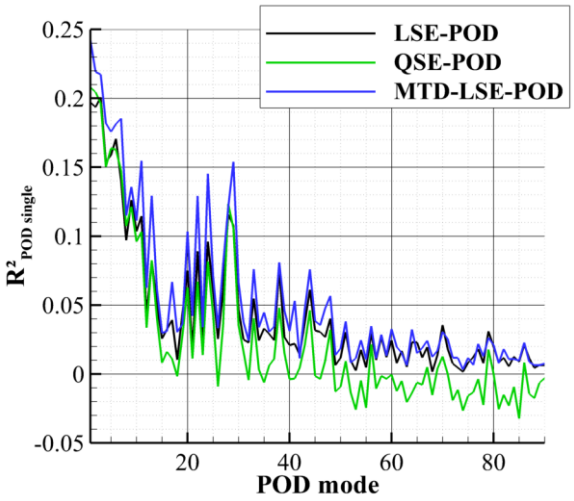


Fig. III.42: Determination coefficient calculated for the first 90 POD modes separately on the validation set using LSE, QSE and MTD-LSE (BFS: low speed PIV)

3.1.3. Results from the 20 and 25 m.s^{-1} database

Determination coefficients from LSE, QSE, LSE-POD and QSE-POD are summarized in Table III.7, Table III.8 for 20 m.s^{-1} upstream velocity case and in Table III.9, Table III.10 for the 25 m.s^{-1} case. Results are perfectly similar to what is obtained when the upstream velocity is 30 m.s^{-1}. Lowering the Reynolds number does not improve the accuracy (in the present range of Reynolds numbers).

		LSE	LSE-POD (626 modes, 90% TKE)	LSE-POD (91 modes, 70% TKE)	LSE-POD (21 modes, 50% TKE)	LSE-POD (6 modes, 32% TKE)
R^2 from original fields	Training	8.3	8.27	8.16	7.46	5.98
	Validation	7.23	7.25	7.2	6.62	5.34
R^2 from POD projected fields	Training	-	9.2	11.6	14.7	18.2
	Validation	-	8.37	10.5	13.4	16.6
R^2_{POD}	Training	-	1.01	4.4	9.83	17.3
	Validation	-	0.69	3.94	9	16.2

Table III.7: Determination coefficients comparison between LSE and several LSE-POD (BFS, U_0 = 20 m.s^{-1}, values are expressed in %)

		QSE	QSE-POD (90%)	QSE-POD (70%)	QSE-POD (50%)	QSE-POD (32%)
R^2 from original fields	Training	11.7	11.2	10.7	9.46	7.39
	Validation	6.75	6.33	6.75	6.54	5.51
R^2 from POD projected fields	Training	-	12.5	15.2	18.7	22.5
	Validation	-	7.3	9.89	13.3	17.2
R^2_{POD}	Training	-	3.38	7.11	13.1	21.1
	Validation	-	-1.82	2.04	7.85	16.1

Table III.8: Determination coefficients comparison between QSE and several QSE-POD (BFS, U_0 = 20 m.s^{-1}, values are expressed in %)

		LSE	LSE-POD (692 modes, 90% TKE)	LSE-POD (99 modes, 70% TKE)	LSE-POD (22 modes, 50% TKE)	LSE-POD (7 modes, 34% TKE)
R^2 from original fields	Training	8.45	8.43	8.32	7.62	6.33
	Validation	8.02	8.03	8	7.43	6.17
R^2 from POD projected fields	Training	-	9.37	11.9	15.2	18.6
	Validation	-	9.26	11.6	14.8	18
R^2_{POD}	Training	-	0.98	4.37	10.2	17.6
	Validation	-	0.67	4.01	9.95	16.8

Table III.9: Determination coefficients comparison between LSE and several LSE-POD (BFS, U_0 = 25 m.s^{-1}, values are expressed in %)

		QSE	QSE-POD (90%)	QSE-POD (70%)	QSE-POD (50%)	QSE-POD (34%)
R^2 from original fields	Training	11.5	11.4	10.9	9.64	7.83
	Validation	6	7	7.42	7.24	6.16
R^2 from POD projected fields	Training	-	12.7	15.5	19.2	23
	Validation	-	8.06	10.8	14.5	18
R^2_{POD}	Training	-	3.37	7.12	13.6	21.6
	Validation	-	-1.88	2.1	9	16.6

Table III.10: Determination coefficients comparison between QSE and several QSE-POD (BFS, U_0 = 25 m.s^{-1}, values are expressed in %)

3.2. Application to the numerical database

LSE was quickly applied to the numerical database. The training was formed of 4 000 snapshots. Among these snapshots 2 300 comes from the data sampled at 20 kHz and 1 700 comes from the data sampled at 250 kHz after they have been downsampled at about 19.3 kHz. The validation set is made of 1060 snapshots from the data at 250 kHz downsampled at 19.3 kHz. In addition, only a subzone of the field is estimated. The subzone goes from –1.3 h to 8.3 h in the streamwise direction and from 0 to 6.5 h in the vertical direction. 17 pressure information at the wall at approximately the same locations than in the experiments are used as conditional events.

The results clearly indicate a strong overfitting due to the bad convergence of the statistical moments used to train the LSE. R^2 is about 13.4% on the training set and -127% on the validation set. The low value obtained in the training set is consistent with the results from the PIV database. Several factors explained that it is slightly higher. First, there is the bad convergence of the statistics. Then, there is the fact that the area estimated (and thus on which R^2 is computed) is different between the experiment and numerical cases. And at last, the small scales of the turbulence are not resolved by the RANS equations (upstream the step and on top of the DES region), thus the velocity field above the detachment zone is better reconstructed on the numerical data than on the PIV data (this will be better addressed in the Chapter IV.3). That is shown in Fig. III.43 where the normalized mean square error for the streamwise velocity estimated on the training set has been plotted for example.

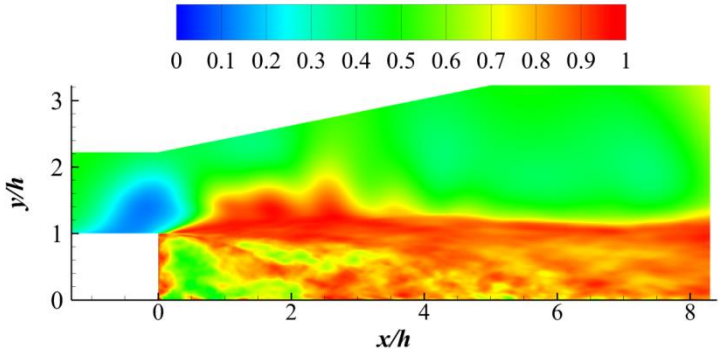


Fig. III.43: Normalized mean square error of the LSE prediction for the streamwise velocity (training set)

3.3. Summary

Using the experimental database, MTD-LSE (using Tikhonov regularization) holds, once more, the best reconstruction and prediction. This time, QSE is shown to suffer from overfitting. If R^2 is increased in the training set, it is decreased in the validation set (even when using Tikhonov regularization). The origin of this overfitting is here the bad convergence of order 3 and 4 statistical moments used to train the QSE. Also, this overfitting is not observed if only Root-Mean-Square quantities are used to evaluate the QSE quality. At last, the use of QSE-POD was shown to present the advantage of allowing disregarding POD modes which prediction penalized the overall estimation due to overfitting. Slight improvement could be achieved this way.

Whatever the SE methods used, the reconstruction and prediction quality is very low compared to what was obtained in the blunt trailing edge case or in the OAT15A airfoil case. Except in a small region downstream the step, above the recirculation, the normalized mean square error is very high (more than 0.9). The bad performances of SE are consistent with the very low levels of correlation between the pressure sensors and the velocity field.

Looking at the estimation of the POD modes separately, it appears that every POD modes are badly estimated (R^2 is lower than 25% whatever the POD mode considered). In addition, the estimation accuracy of the POD modes rapidly decreases when POD mode rank increases.

4. Flow around a wall mounted cube case

In this part, only the case where the upstream velocity is 30 $m.s^{-1}$ is discussed to illustrate the performance of Stochastic Estimation in the situation with the highest Reynolds number. Determination coefficients for the other cases are summarized in Annex H.4.

4.1. Application to the low speed PIV database

As considered in Chapter II.3.3.4.1, the training set is formed of the data from 2 PIV runs. It contains 6 800 PIV snapshots. The validation set is made of the remaining PIV run and contains 3 400 PIV snapshots.

4.1.1. Single-time Stochastic Estimation

4.1.1.1. Linear Stochastic Estimation (LSE) and Quadratic Stochastic Estimation (QSE)

- *Vertical plane:*

The average correlation module between the 23 pressure sensors and the velocity field is displayed in Fig. III.44. Levels of correlation are similar to the BFS case. The vertical velocity component appears to be less correlated with the pressure sensors than the streamwise velocity component, with correlation levels particularly low (under 0.06) downstream the step for $y/h < 1$. For both streamwise and vertical velocity component, the highest correlation levels are found above the cube and downstream of the cube for $y/h > 1.5$. In addition, for the streamwise velocity, the correlation levels are also higher downstream of the cube around the reattachment point. Concerning the spanwise velocity component, the maximum of correlation level is lower than for the other two components. But the correlation is more uniformly spread over the entire field. The lowest correlation is found directly above the cube, in a similar fashion as for the other two components. The highest correlations are found downstream of the step, near the wall, around the reattachment location. These low levels of correlation show that the LSE is likely to be inaccurate as for the BFS case. The vertical velocity is also likely to be predicted with the lowest accuracy.

An example of an instantaneous field (from the validation set) and its estimation is shown in Fig. III.45 (streamwise velocity), Fig. III.46 (vertical velocity) and Fig. III.47 (spanwise velocity). Similarly to the BFS case, scales of the predicted field and the original field are not the same and the velocity fluctuation levels predicted are lower than the original one. In addition, the predicted fields present only large scale structures of low fluctuation levels whereas the original field contains a large variety of small scale structures of high fluctuation levels. However, some qualitative similitudes between the original and predicted fields can be observed. For the streamwise velocity, the area of positive and negative fluctuations just above the cube seems correctly predicted. The large negative fluctuation area predicted downstream of the cube seems to be underlying in the original field. For the vertical velocity, the positive fluctuation area at about $x/h = 3$ is predicted. At last, for the spanwise velocity, the negative fluctuation region spanning from $x/h = 2$ to 4 seems to be well predicted. Nevertheless, the predicted field and the original one remains quite different. The prediction quality appears to be poor. It is confirmed by the determination coefficient that is about 14.5%. The reconstruction quality is very similar with R^2 equals to 14.7%.

The normalized mean square error (calculated on the validation set) is shown in Fig. III.48. The link between the average correlation module and the normalized mean square error can once more be appreciated. The regions of lowest error correspond to the ones of highest correlations. In overall the error remains high in the entire field with values higher than 0.5. From this figure, the spanwise velocity seems to be better predicted than the others and the vertical velocity the less. That is confirmed by calculating the determination coefficient for each velocity component separately. R^2 is 11.5% for the streamwise velocity, 4.8% for the vertical velocity and 23% for the spanwise velocity.

The prediction by QSE of the same instantaneous flow field as in Fig. III.45, Fig. III.46 and Fig. III.47 is shown in Fig. III.49. As for the LSE prediction, the scales are different than for the original field. The extrema are of lower amplitudes than the original ones, but of higher amplitude than the LSE prediction ones. As observed in the BFS case, the structures predicted by QSE are of slightly higher fluctuation levels. The global shape of the predicted structures does not seem particularly changed between the LSE and QSE case, but smaller structures are predicted in QSE than in LSE. For instance, the negative streamwise velocity fluctuation region predicted by LSE just above the cube is split into two structures in the QSE prediction. Qualitatively it remains difficult to judge if the modifications observed in the QSE leads to a better prediction than in LSE. The normalized mean square error of the QSE prediction is plotted in Fig. III.50. In this figure, the error seems to mainly increase between the LSE and QSE case. In addition, the error from the QSE and LSE are also very similar (areas of lowest and highest error remain the same). Therefore, the QSE does not

hold any important improvement of prediction accuracy. Actually, from the determination coefficient evaluation, it appears that the prediction accuracy of the QSE is lower than the one of the LSE. Indeed, R^2 is only 13.4%. As in the BFS case, it is due to the fact that QSE suffers from overfitting in this case. The training set R^2 is 20.8% which is much higher than for the validation set. In the present situation, using QSE leads to a more complicated model than LSE. QSE holds 276 coefficients, instead of 23. In addition, the convergence study (see Chapter II.3.3.4.1) showed that the statistical moments used to determine the QSE coefficients were not well converged. Thus the QSE model does not represent the entire phenomena leading to a deterioration of the prediction. The VIF goes from 1.3 to 1.9 between the LSE and QSE thus the main problem is probably the lack of convergence of the data on the training set. At last, the use of Tikhonov regularization shows only a slight improvement (less than 0.5% on the validation set R^2).

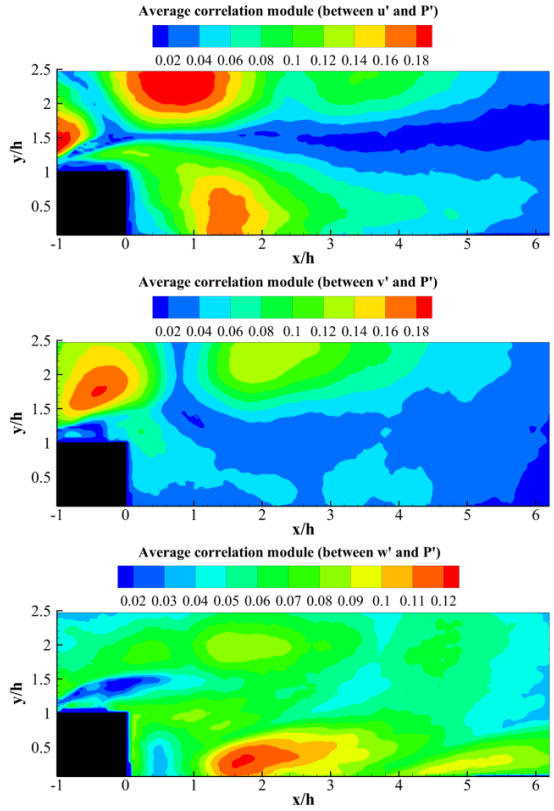


Fig. III.44: Average correlation module between the streamwise velocity (top), the vertical velocity (middle) or the spanwise velocity (bottom) and the set of 23 pressure sensors

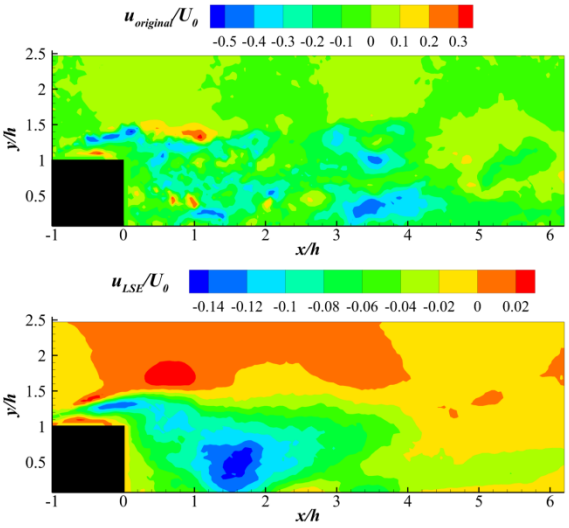


Fig. III.45: Instantaneous fluctuating streamwise velocity field. Top: original, bottom: LSE prediction.

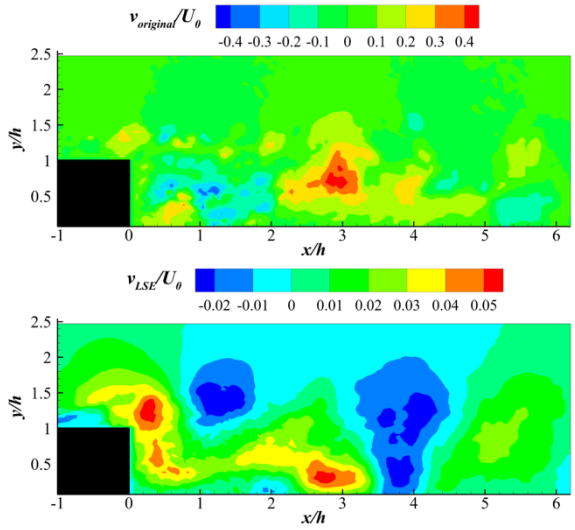


Fig. III.46: Instantaneous fluctuating vertical velocity field. Top: original, bottom: LSE prediction.

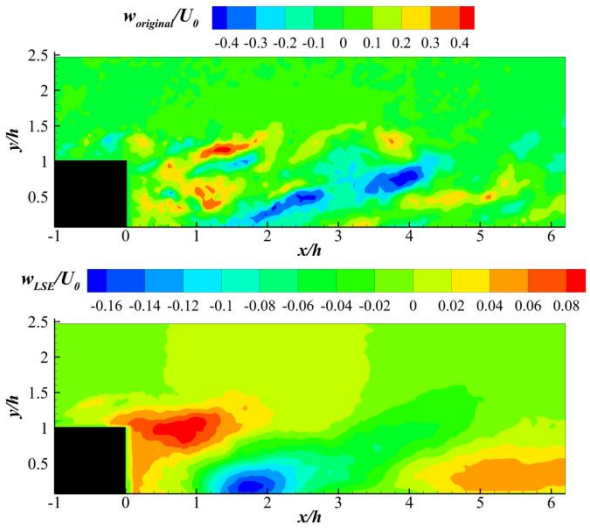


Fig. III.47: Instantaneous fluctuating spanwise velocity field. Top: original, bottom: LSE prediction.

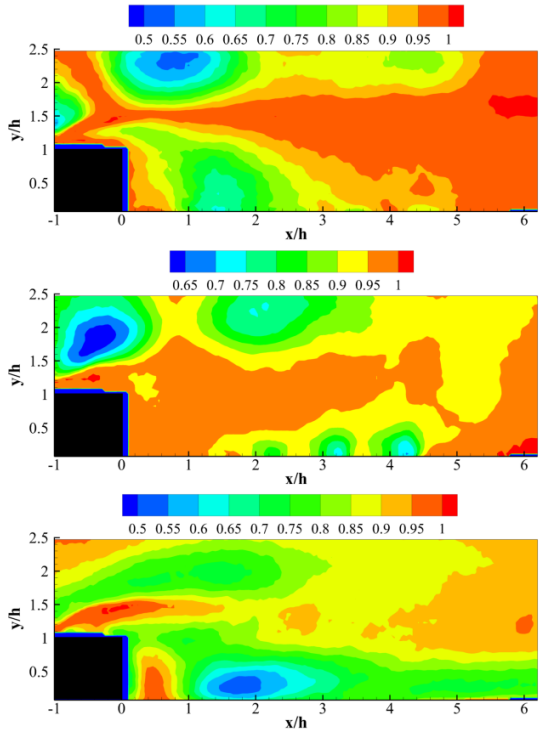


Fig. III.48: Normalized mean square error of the LSE prediction for the streamwise velocity (top), the vertical velocity (middle) and the spanwise velocity (bottom)

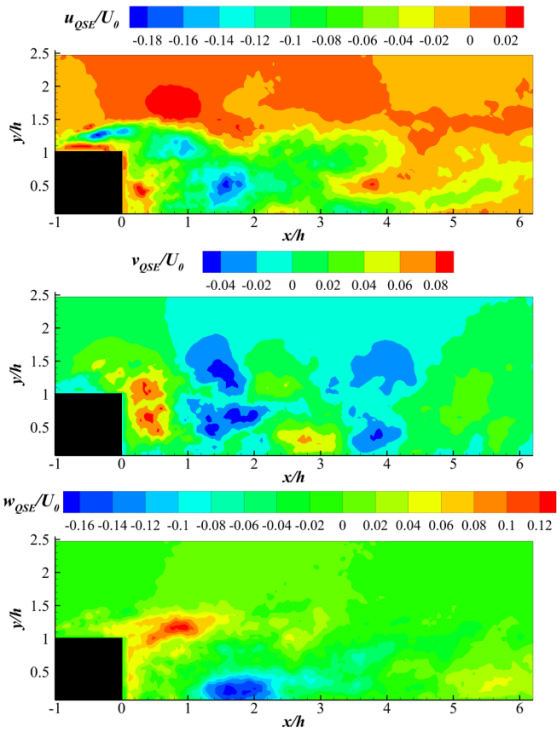


Fig. III.49: Instantaneous fluctuating velocity field predicted by QSE. Top: streamwise component; middle: vertical component; bottom: spanwise component

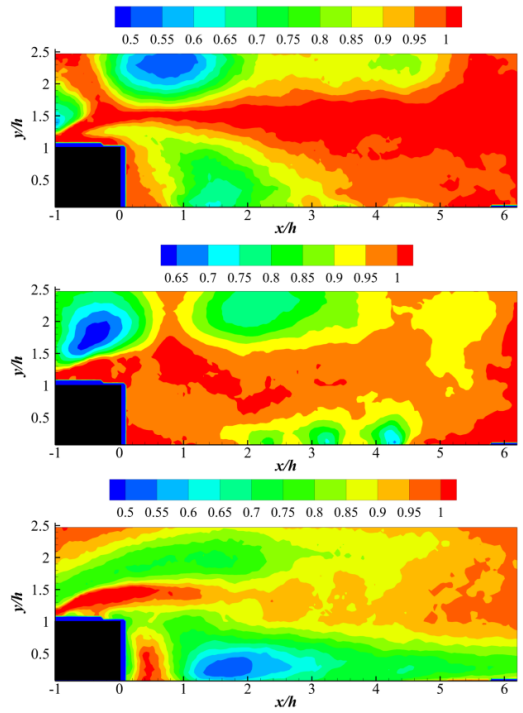


Fig. III.50: Normalized mean square error of the QSE prediction for the streamwise velocity (top), the vertical velocity (middle) and the spanwise velocity (bottom)

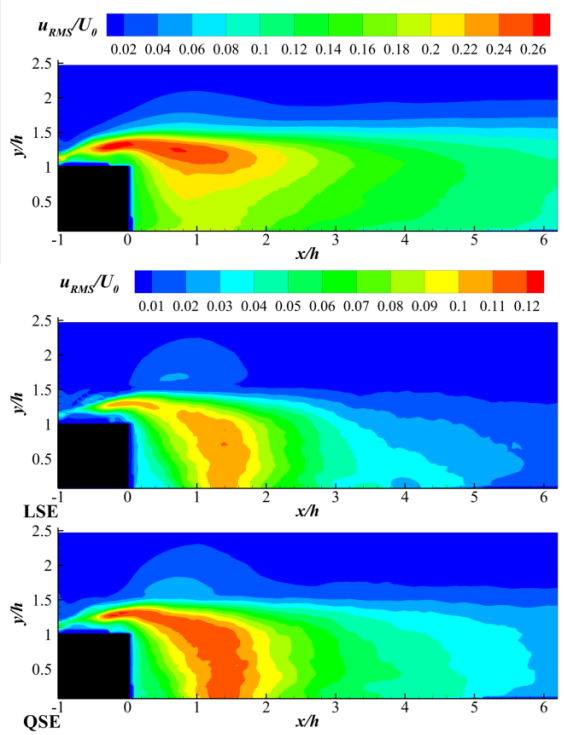


Fig. III.51: Streamwise root-mean-square velocity; top: original; middle: LSE estimation; bottom: QSE estimation (validation set)

As done in the BFS case, the estimation of the velocity RMS is evaluated. Velocity RMS fields are displayed in Fig. III.51. Only the streamwise velocity component is shown. As expected from the observations made in the BFS case, RMS velocity is better predicted by QSE than by LSE even if the QSE suffers from overfitting. It confirms that the evaluation of RMS quantities prediction gives no information about the prediction quality of instantaneous quantities.

- *Horizontal plane:*

The average correlation module between the 23 pressure sensors and the velocity field is displayed in Fig. III.52. From these figures, some regions appear more correlated than others. And the LSE is expected to be more accurate just downstream of the cube for the prediction of the streamwise velocity, and around $x/h = 2$ for the spanwise velocity. Concerning the vertical velocity, the correlation levels are lower than for the two other velocity components. The correlation level is around 0.5 for most of the field. Thus, one can expect this component to be predicted with the worst accuracy between the three velocity components.

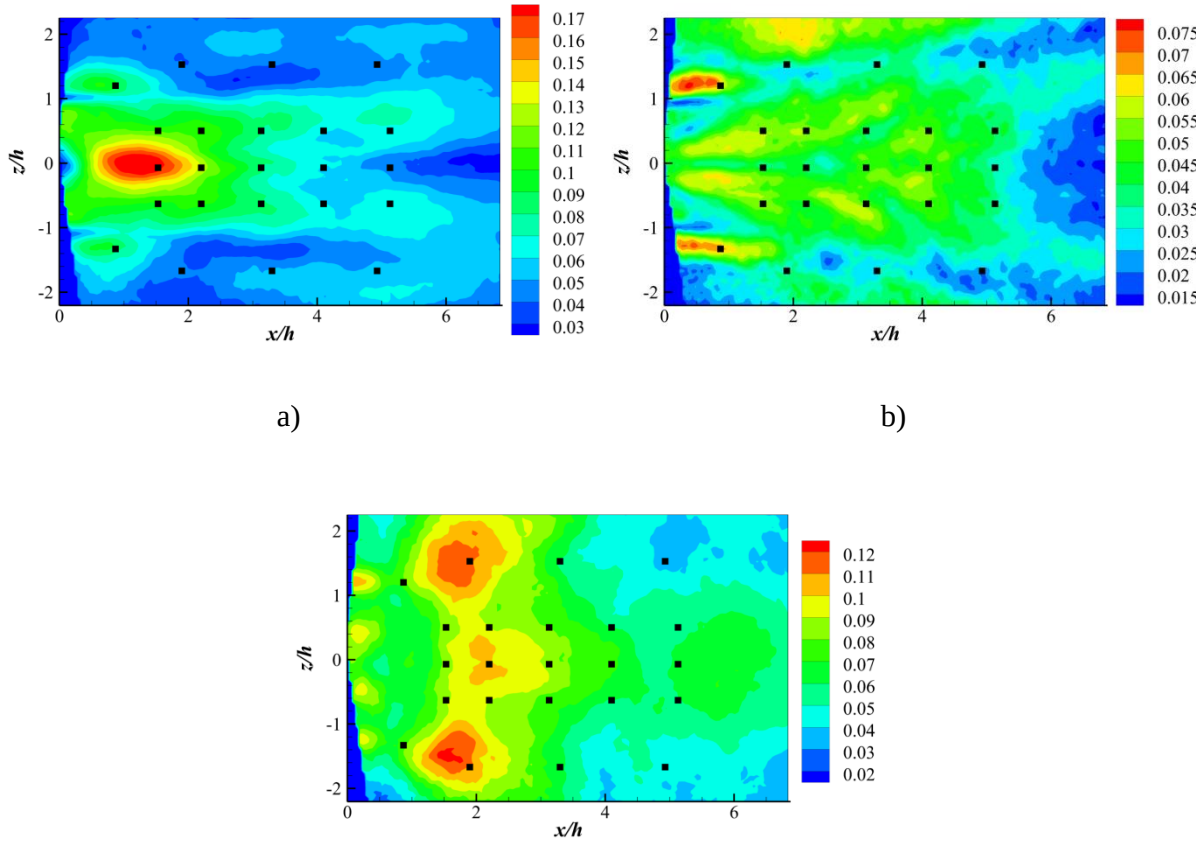
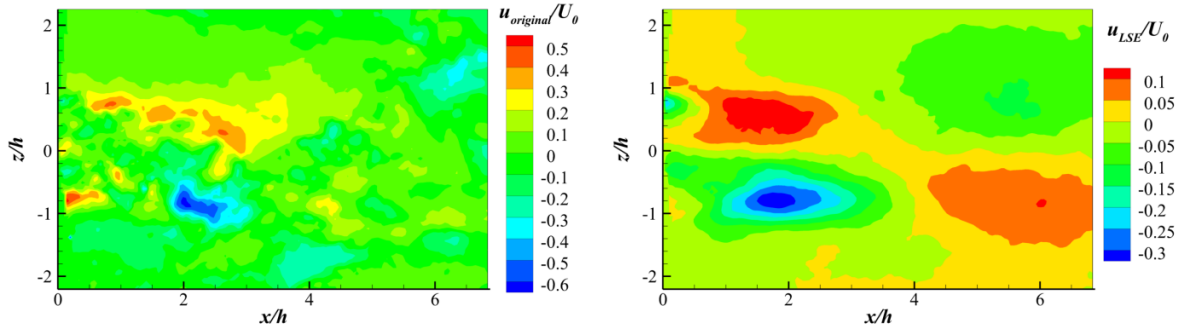


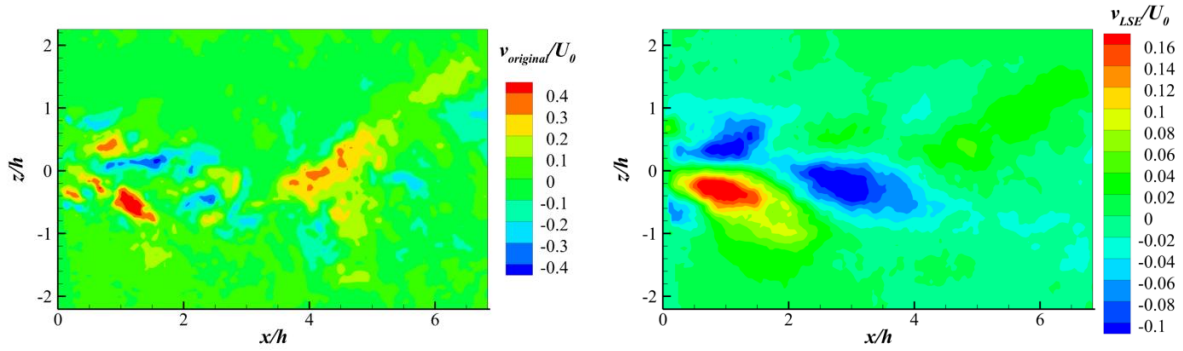
Fig. III.52: Average correlation module between the streamwise velocity (a), the vertical velocity (b) or the spanwise velocity (c) and the set of 23 pressure sensors (black squares)

The estimation of an instantaneous velocity field, from the validation set, by LSE is shown in Fig. III.53, Fig. III.54 and Fig. III.55. Once more the predicted levels of velocity fluctuations are lower than the original ones (scales are not the same between the original field and the predicted one). However, the predicted extrema are only about 2 times lower than the originals. In addition, the predicted field seems qualitatively closer to the original one than it is the case for the BFS or the vertical plane. The predicted structures are larger than the original ones but their locations and shapes are somehow in agreement with the original structures. For instance, the succession of negative to positive streamwise velocity fluctuations at $x/h \approx 2$, when z/h increases, is correctly predicted. For the spanwise velocity, the succession, when x/h increases, is also correctly predicted. For the vertical velocity, the agreement is not as clear. For instance, the large area of negative velocity between $x/h = 2$ and 4, is not present in the original field. Nevertheless, for $x/h < 2$, the negative and positive predicted structures can be linked to the main structures of the original field. These similitudes between the predicted and original field is not a particularity of the instantaneous field shown here. And the LSE prediction seems in this case to qualitatively correspond to the general trends of the original field.



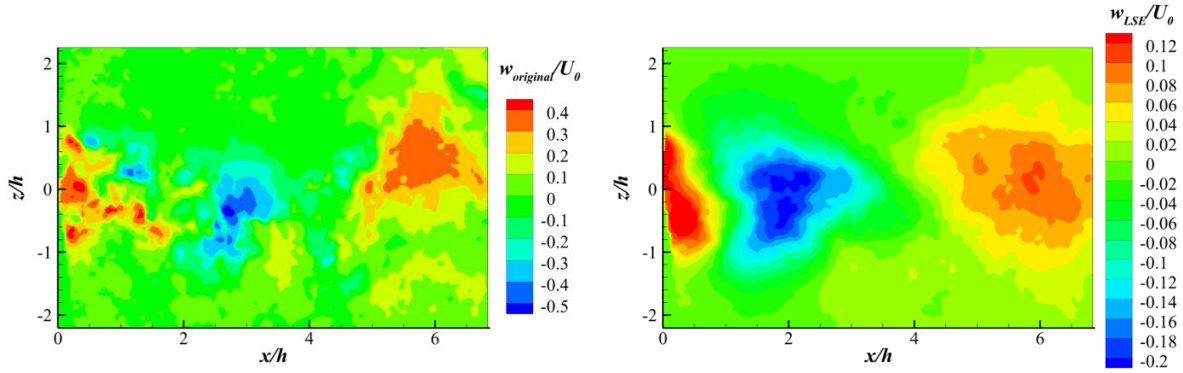
a) b)

Fig. III.53: Instantaneous streamwise velocity field from the validation set. a) original field; b) LSE prediction.



a) b)

Fig. III.54: Instantaneous vertical velocity field from the validation set. a) original field; b) LSE prediction.

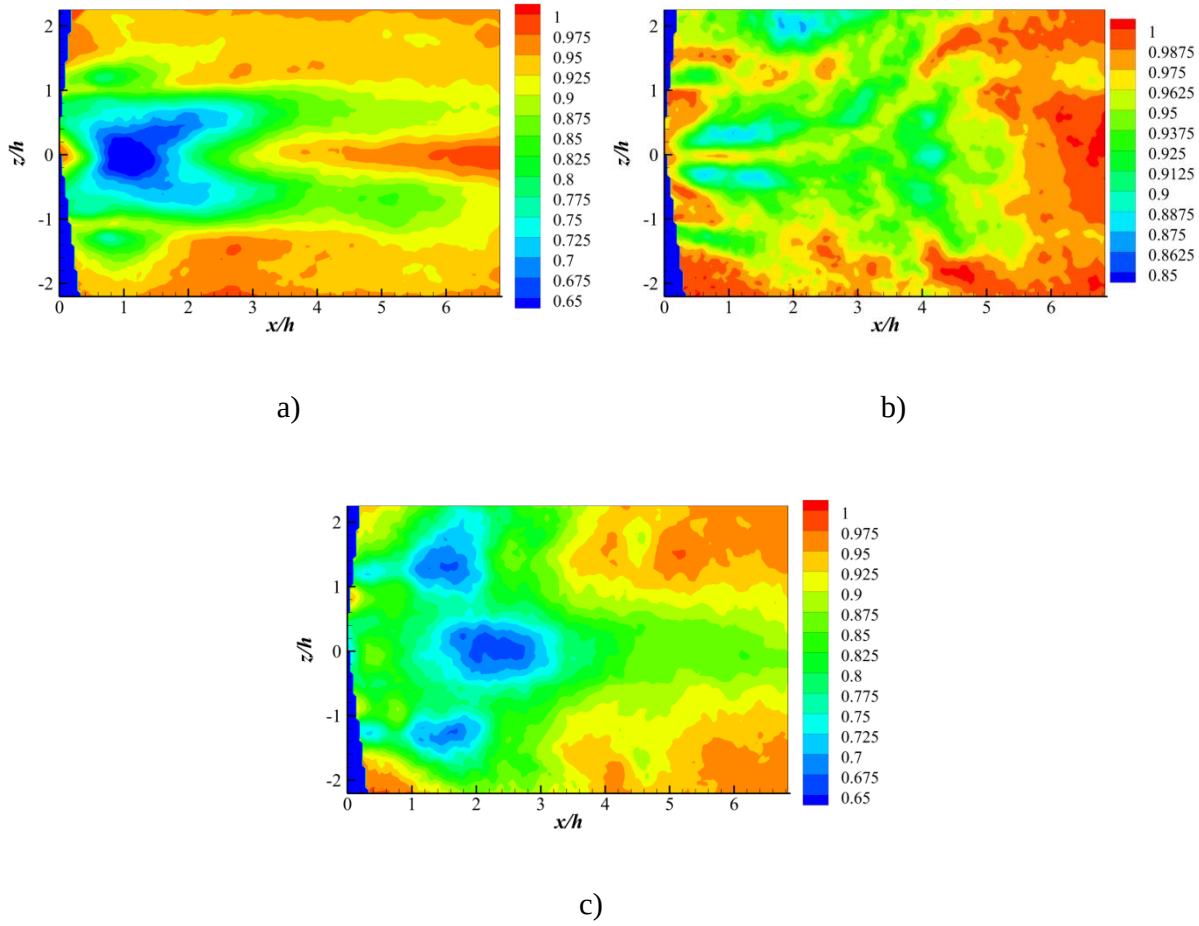


a) b)

Fig. III.55: Instantaneous spanwise velocity field from the validation set. a) original field; b) LSE prediction.

The determination coefficient is, however, not especially higher than for the vertical plane. It is 16.7% on the training set and 15.7% on the validation set. The normalized mean square error is

plotted in Fig. III.56. The structures formed by the error are once more similar to those observed on the average correlation module. In overall, the error is slightly lower than for the vertical plane (mainly for the streamwise and spanwise velocity). That may come from the fact that in overall it is closer to the sensors located at the wall. However, the error remains in general quite high. In particular, the error on the vertical velocity prediction is very high. Indeed, the error is higher than 0.9 for almost the entire flow field. The difference of estimation quality between the three velocity components is perfectly seen on the determination coefficient calculated for each component separately: 17.5% (18.1%) for the streamwise component, 5.53% (6.1%) for the vertical component and 19.2% (20%) for the spanwise component (values in parentheses are from the training set).



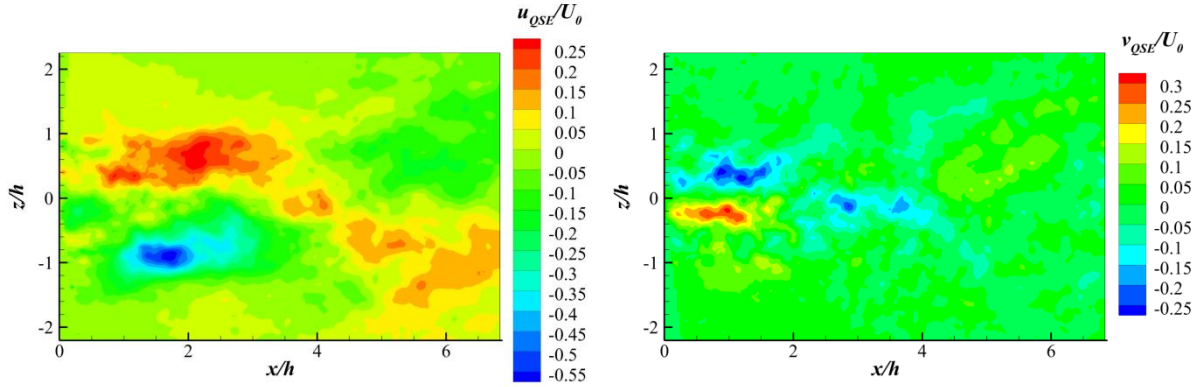
a) b)

c)

Fig. III.56: Normalized mean square error of the LSE prediction for the streamwise velocity (a), the vertical velocity (b) and the spanwise velocity (c)

QSE prediction of the instantaneous velocity field is shown in Fig. III.58. The levels of fluctuations are very close to the original ones (especially for the spanwise velocity), but are still slightly lower. In overall, locations of large negative or positive velocity remain unchanged between the LSE prediction and the QSE prediction and the levels of fluctuations are mostly increased. The QSE prediction also holds smaller structures than the LSE prediction. If qualitatively the QSE prediction could look more accurate than the LSE, the determination coefficient shows that QSE suffers, once more of overfitting, and its overall prediction accuracy is lower than the LSE one. For QSE, R^2 is 23.4% on the training set and 14.7% for the validation set. As for the vertical plane, the overfitting comes from the bad convergence of third and fourth statistical moments used to train the QSE. The use of Tikhonov regularization does not particularly improve the accuracy. R^2 increases by less than 0.1%. The deterioration of the prediction can also be seen on the normalized mean square error (see

annex H.4.1, Fig. H.6). Concerning the prediction of the velocity RMS, results are shown in Fig. III.59 for the streamwise velocity and are consistent with previous conclusions. QSE improves the prediction of the velocity RMS even if it deteriorates (in average) the prediction of the instantaneous velocity fields of the validation set.



a) b)

Fig. III.57: Instantaneous fluctuating velocity field predicted by QSE. a) streamwise velocity, b) vertical velocity

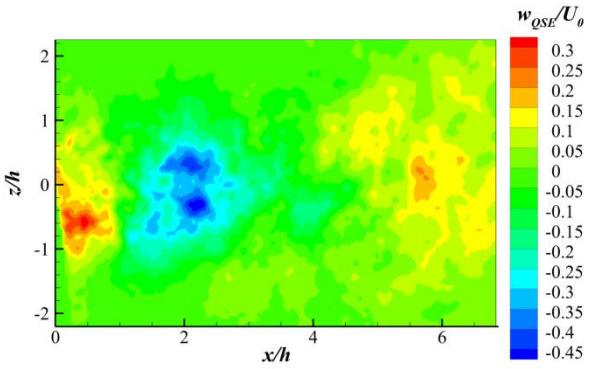


Fig. III.58: Instantaneous fluctuating velocity field predicted by QSE (spanwise velocity)

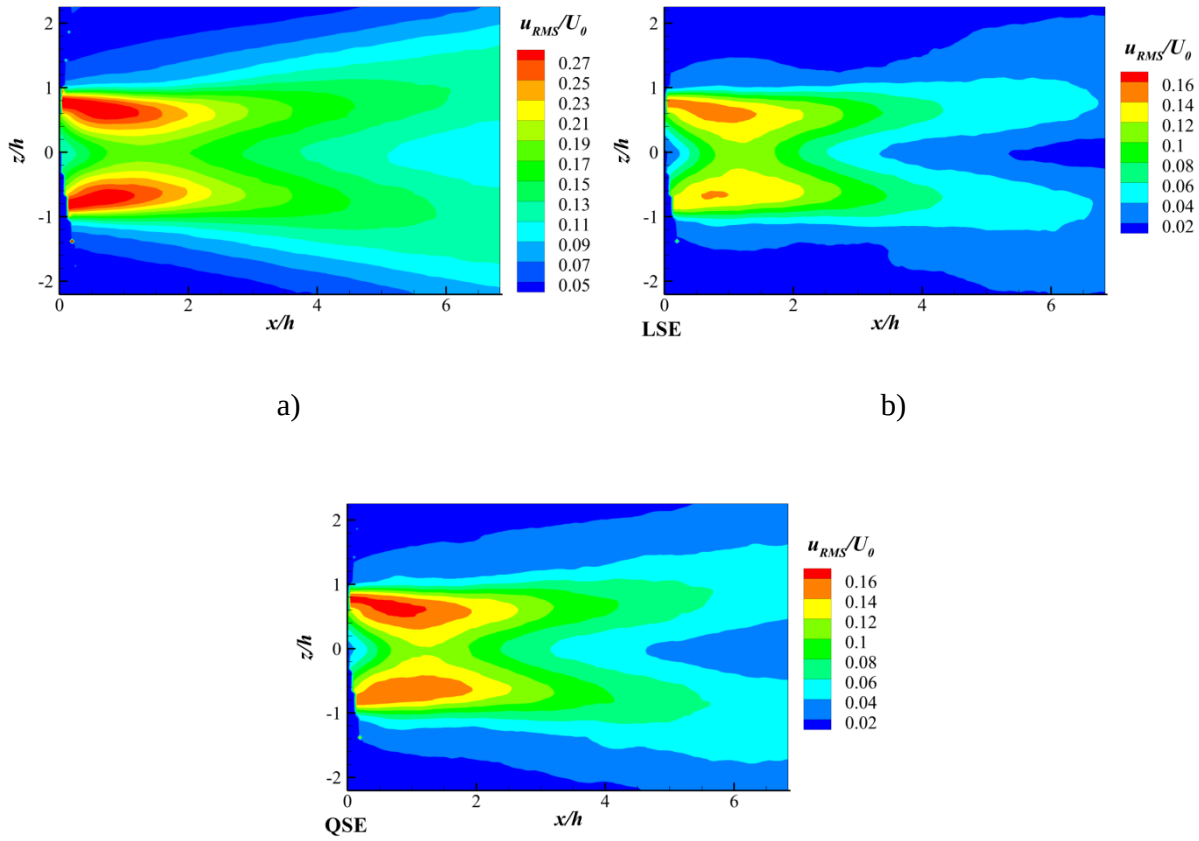


Fig. III.59: Streamwise root-mean-square velocity. a) original, b) LSE estimation, c) QSE estimation

4.1.1.2. Modified Stochastic Estimation (SE-POD)

- *Vertical plane:*

The POD basis is computed using the three velocity components and data from the training set only. This spectrum is similar to the one obtained in the BFS case and shown in Fig. III.60. 25 POD modes are necessary to retain 50% of the TKE and 507 to retain 90% of the TKE. The flow dynamics is not captured with a few number of POD modes. However, the first three POD modes contain more energy and together they represent more than 23% of the TKE.

Prediction of the spanwise velocity field using LSE-POD is plotted in Fig. III.61. The same instantaneous field as in Fig. III.55 is considered. The POD basis is formed of the first 94 POD modes which contain 70% of TKE. Once again, the predicted structures are very similar to those obtained with a direct LSE. In addition, the fluctuation levels are unchanged (the same is observed for the two other velocity components).

The determination coefficients of 4 different LSE-POD estimations are compared with the LSE estimation in Table III.11. The results are in perfect agreement with the observations made in the BFS case. Using LSE-POD does not improve the reconstruction accuracy and most of the time the prediction accuracy too. However, when the POD basis retains 70% or 90% of the TKE, R^2 on the validation set slightly increases. That indicates that POD modes of highest rank suffer the most

from overfitting (their prediction actually deteriorate the prediction of the entire flow field). Evolutions of R^2 calculated with the POD projection of the original field as reference and of R^2 directly calculated from the POD coefficients are in agreement with previous results. They decrease when more POD modes are considered. Thus, POD modes of higher ranks are estimated with a globally decreasing accuracy.

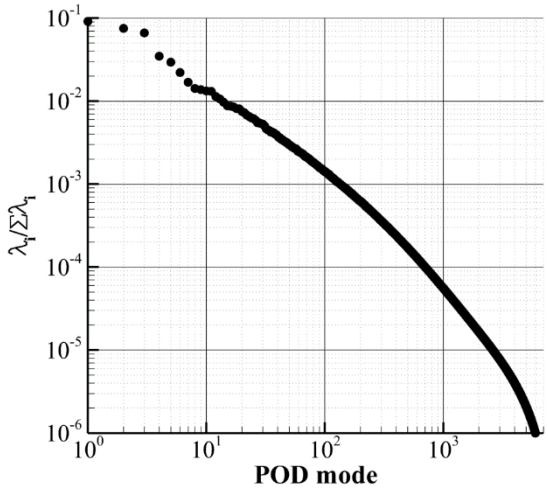


Fig. III.60: Energy spectrum of all POD modes (Cube, vertical plane, low speed PIV)

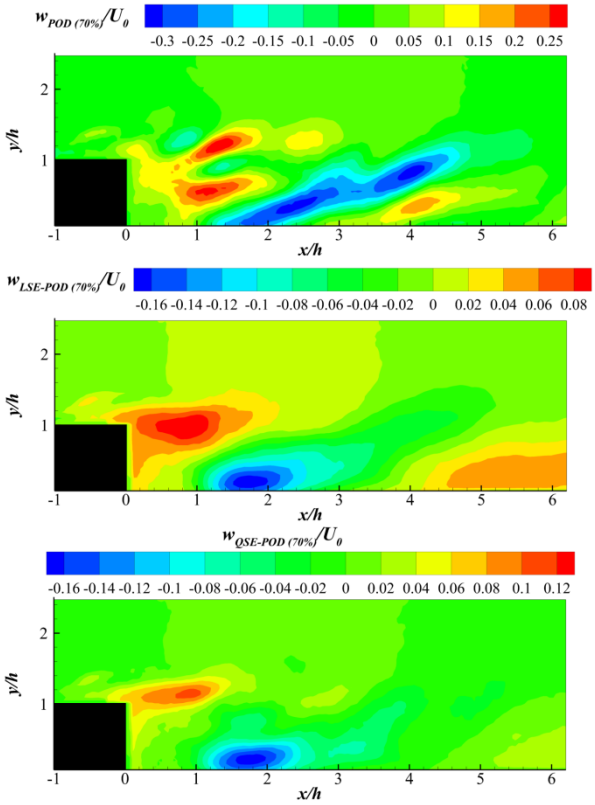


Fig. III.61: Instantaneous fluctuating spanwise velocity field. Top: POD projection; middle: LSE-POD prediction; bottom: QSE-POD prediction

$R^2_{\text{POD Single}}$ of the prediction of the first 500 POD modes by LSE is plotted in Fig. III.62. The figure shows that the first three POD modes are far better predicted (it is also true for the reconstruction) than any other POD modes. For each of these three POD modes $R^2_{\text{POD Single}}$ is higher than 48%, whereas it is lower than 12% for any other POD modes. That may explain why the prediction seems to be qualitatively correct while the R^2 remains low. Some information about the velocity field is predicted with a relatively high accuracy. This is a strong difference with the BFS case. The frequency content of these POD modes is described in Chapter IV.5.2 and gives some insight on the reason for this difference.

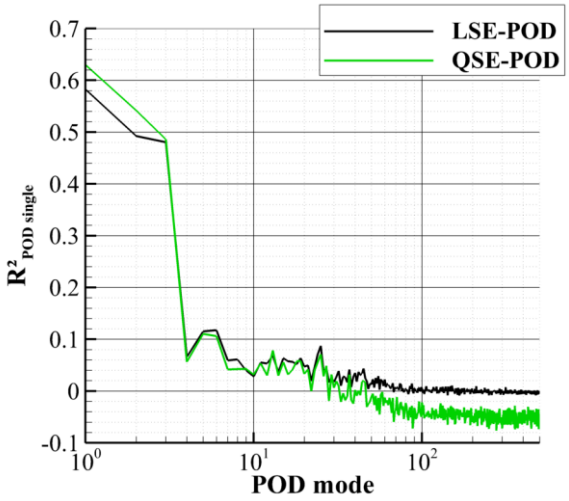


Fig. III.62: Determination coefficient calculated for each POD mode separately on the validation set using LSE and QSE (Cube: low speed PIV, vertical plane)

Results obtained with QSE-POD are in agreement with observations made in the BFS case. The QSE-POD estimated fields are very similar to the QSE estimated ones (see Fig. III.61 and Fig. III.49). Determination coefficients are detailed in Table III.12. QSE clearly improves the reconstruction accuracy, but the prediction accuracy is deteriorated. The use of POD is able to compensate the overfitting from the QSE (as observed in the BFS case). Indeed, when only 25 POD modes (50% TKE) are used, R^2 reaches 15% which is of the same order of magnitude than what is obtained with LSE. That is coherent with what can be observed in Fig. III.62. Indeed, $R^2_{POD\ Single}$ of most of POD modes of rank higher than 30 is lower than 0. Thus, due to the overfitting, their prediction deteriorates the prediction of the entire velocity field by QSE. Nevertheless, the improvement is still marginal (only 0.5%). Actually, the QSE improves mainly the prediction of the first two POD modes.

		LSE	LSE-POD (90%, 507 modes)	LSE-POD (70%, 94 modes)	LSE-POD (50%, 25 modes)	LSE-POD (33%, 7 modes)
R^2 from original fields	Training	14.7	14.7	14.6	14.2	13.2
	Validation	14.5	14.7	14.7	14.4	13.5
R^2 from POD projected fields	Training	-	16.4	20.8	28.1	39.3
	Validation	-	16.7	21.2	28.4	39.5
R^2_{POD}	Training	-	1.2	4.49	11.9	27.3
	Validation	-	0.72	4.22	11.8	27.8

Table III.11: Determination coefficients comparison between LSE and several LSE-POD (Cube: low speed PIV, vertical plane; values are expressed in %)

		QSE	QSE-POD (90%)	QSE-POD (70%)	QSE-POD (50%)	QSE-POD (33%)
R^2 from original fields	Training	20.8	20.4	19.4	17.9	15.8
	Validation	13.4	13.9	14.7	15	14.3
R^2 from POD projected fields	Training	-	22.7	27.7	35.5	47
	Validation	-	15.8	21.3	29.7	41.8
R^2_{POD}	Training	-	5.72	10.1	19	35.1
	Validation	-	-3.66	1.59	11.4	28.7

Table III.12: Determination coefficients comparison between QSE and several QSE-POD (Cube: low speed PIV; values are expressed in %)

- *Horizontal plane:*

The POD energy spectrum is shown in Fig. III.63. Compared to the vertical plane POD, it is the first two POD modes that contain really more energy than the others. They contain about 25% of the TKE. Nevertheless, to capture a significant amount of the TKE, a large number of POD modes is necessary. 27 POD modes are required to retain 50% of TKE and 660 to retain 90%.

Prediction of the streamwise velocity field shown in Fig. III.53 by LSE-POD is plotted in Fig. III.65. The POD projection is shown in Fig. III.64. The LSE-POD predicted field and the LSE one are very similar (it is also true for the other velocity components not shown here). As previously observed, the use of LSE-POD does not change the location and shape of the estimated structures.

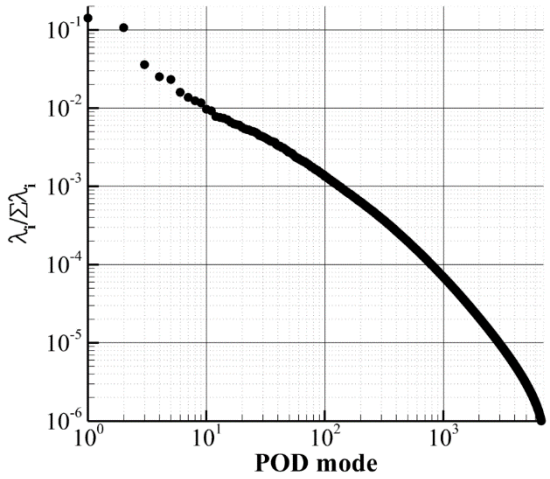


Fig. III.63: Energy spectrum of all POD modes (Cube, horizontal plane, low speed PIV)

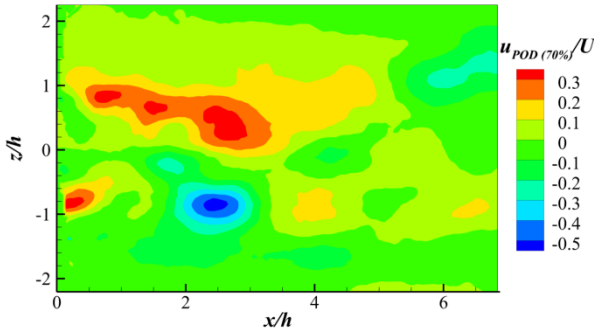


Fig. III.64: POD projection of an instantaneous streamwise velocity field from the validation set (POD basis contains 70% of TKE)

R^2 of several LSE-POD are summarized in Table III.13 and can be compared with LSE values. Evolutions of R^2 with the part of TKE contained by the POD basis are in agreement with the observations made so far. Since the LSE does not suffer from any particular overfitting, the use of POD does not improve the prediction accuracy. The POD modes seem predicted (and reconstructed) with lower and lower accuracy when their rank increases. That is confirmed when looking at $R^2_{POD\ Single}$ in Fig. III.66. As for the vertical plane, the first three POD modes are far better predicted than the others ($R^2 > 45\%$ instead of $R^2 < 14\%$). But surprisingly, it is the third POD mode that is predicted with the highest accuracy. This is further investigated in Chapter IV.5.

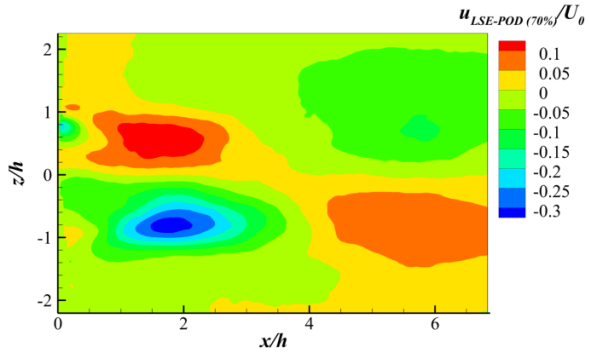


Fig. III.65: LSE-POD prediction of an instantaneous streamwise velocity field

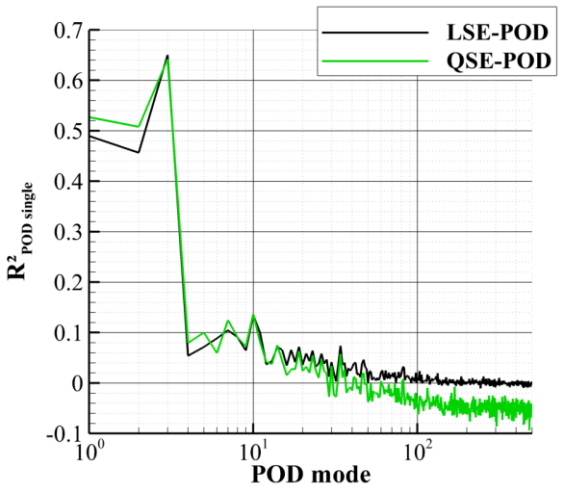


Fig. III.66: Determination coefficient calculated for each POD mode separately on the validation set using LSE and QSE (Cube: low speed PIV, horizontal plane)

		LSE	LSE-POD (90%, 660 modes)	LSE-POD (70%, 117 modes)	LSE-POD (50%, 27 modes)	LSE-POD (33%, 5 modes)
R^2 from original fields	Training	16.7	16.7	16.5	16	14.8
	Validation	15.7	15.7	15.7	15.3	14.1
R^2 from POD projected fields	Training	-	18.5	23.6	31.9	44.3
	Validation	-	18	22.9	30.9	43.3
R^2_{POD}	Training	-	1.23	4.45	11.9	35.1
	Validation	-	0.72	3.94	11.5	34.2

Table III.13: Determination coefficients comparison between LSE and several LSE-POD (Cube: low speed PIV, horizontal plane; values are expressed in %)

QSE-POD results are also in agreement with observations from the BFS case. QSE-POD estimations are very similar to QSE estimations (see Fig. III.67 for instance). From the determination coefficients displayed in Table III.14, one can see that QSE-POD also suffers from overfitting. But using POD subbases can compensate the overfitting. Thus, when 27 POD modes only are considered, R^2 reaches 16.3% for the validation set which is slightly higher than in the LSE case. As for the vertical plane prediction, it appears in Fig. III.66 that POD modes of ranks higher than 30 predictions are very inaccurate. It is then better not to predict them by QSE. In overall, the QSE only improves the prediction of the first two POD modes by a small percent.

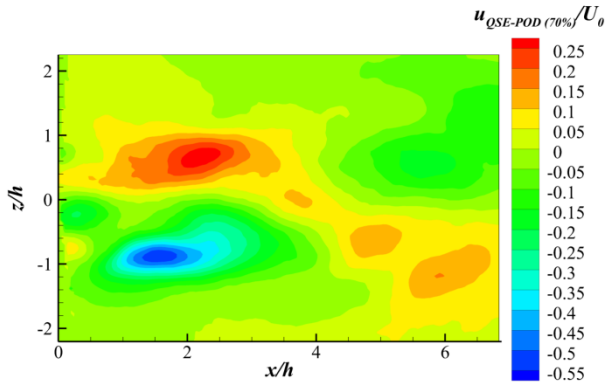


Fig. III.67: QSE-POD prediction of an instantaneous streamwise velocity field

		QSE	QSE-POD (90%, 660 modes)	QSE-POD (70%, 117 modes)	QSE-POD (50%, 27 modes)	QSE-POD (33%, 5 modes)
R^2 from original fields	Training	23.4	23	21.9	20.3	17.8
	Validation	14.7	15.1	16	16.3	15.3
R^2 from POD projected fields	Training	-	25.5	31.3	40.4	53.5
	Validation	-	17.4	23.4	32.9	46.8
R^2_{POD}	Training	-	5.93	10.3	19.3	44
	Validation	-	-3.87	1.08	11.1	36.9

Table III.14: Determination coefficients comparison between QSE and several QSE-POD (Cube: low speed PIV, horizontal plane; values are expressed in %)

4.1.2. Multi-Time-Delay Stochastic Estimation (MTD-SE)

A perfectly similar study as for the backward facing step (BFS) is conducted where several sets of delays (with constant time step between delays) are tested.

- *Vertical plane:*

The evolution of R^2 (calculated on the validation set) as function of the number of delays used for several constant time steps between consecutive delays is plotted in Fig. III.68. These values are obtained using Tikhonov regularization (without Tikhonov regularization, R^2 decreases almost immediately when the number of delays increases). The regularization parameter is obtained through cross-validation [72]. As expected, an optimal time window is observed for the choice of delays ranging from 0 to 0.7-1 $T_{shedding}$ (St = 0.1 is considered as the shedding frequency). Concerning the reconstruction, R^2 increases as the number of delays increases (without Tikhonov regularization) and reaches 64% when 150 delays taken every 10^{-2} $T_{shedding}$ are used.

The highest R^2 (for the validation set) achieved using MTD-LSE is 23.6% which corresponds to an increment by about 9% compared to the LSE. Therefore, MTD-LSE also holds the best prediction accuracy in this case (between LSE, QSE and MTD-LSE). However, the global accuracy remains low. From these curves, it appears that reducing the time step between the delays first improves the prediction accuracy. But the curves obtained for $\Delta\tau = 10^{-2}$ $T_{shedding}$ and $\Delta\tau = 2\times10^{-2}$ $T_{shedding}$ are very close. Thus it seems that are a limit on the time step under which no improvement is obtained on the prediction accuracy of the MTD-LSE. This was also observed in the BFS case.

The prediction by MTD-LSE of the streamwise velocity of the same instantaneous field as in Fig. III.45 is plotted in Fig. III.70. The predicted structures are quite similar to those obtained by LSE and QSE. Smaller structures are predicted by MTD-LSE than by LSE, but not particularly than QSE.

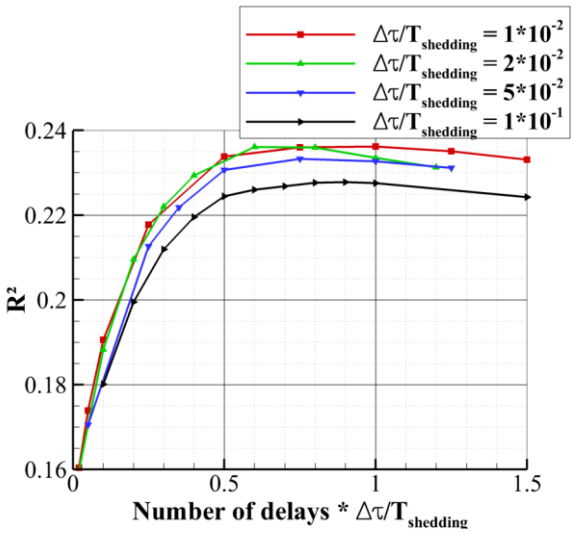


Fig. III.68: Determination coefficient as function of the number of delays for MTD-LSE estimations of the validation set (Low speed PIV cube experiment, vertical plane)

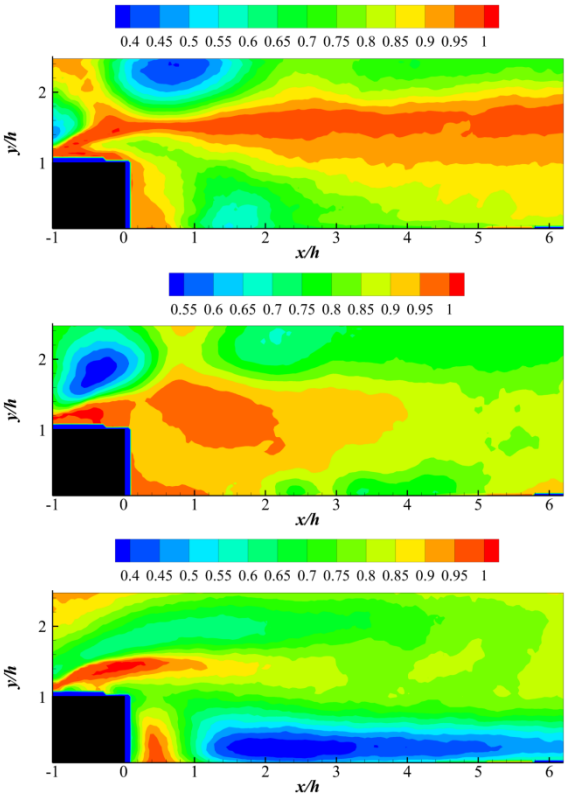


Fig. III.69: Normalized mean square error of the MTD-LSE prediction for the streamwise velocity (top), the vertical velocity (middle) and the spanwise velocity (bottom) (100 delays every 10^{-2} $T_{shedding}$)

The normalized mean square error for the three velocity components are plotted in Fig. III.69. The figure clearly shows that the error decreases almost everywhere in the field and for every velocity component compared to LSE and QSE. Improvements seem particularly important downstream the reattachment location. The error is not decreased as much in the recirculation and just above the cube as it is in other areas. That can be particularly observed for the spanwise velocity. The high error regions, just above and just downstream of the step in the LSE prediction, are also high error regions when using MTD-LSE.

At last, MTD-LSE-POD is investigated. Determination coefficient of the prediction of POD mode separately using the best MTD-LSE-POD situation is plotted in Fig. III.71 (R^2 from LSE and QSE are also shown). From this figure, the improvement of prediction accuracy of MTD-LSE-POD over LSE and QSE appears to mainly concern the first two POD modes. R^2 increases by more than 30% and reaches values higher than 85% for both POD modes. POD modes of rank up to 70 are also better predicted by MTD-LSE-POD than by the other methods. But POD modes of higher rank are not recovered by MTD-LSE-POD. For these POD modes all the tested SE methods fails to correctly predict their temporal evolution.

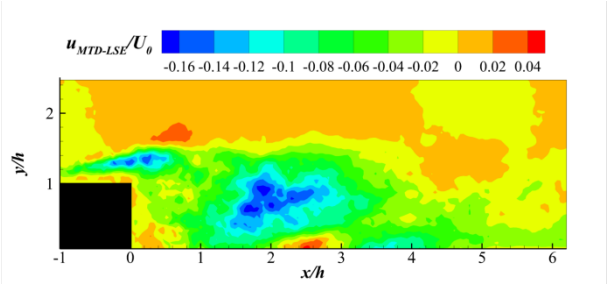


Fig. III.70: Instantaneous streamwise velocity predicted by MTD-LSE (using 100 delays every 10^{-2} $T_{shedding}$)

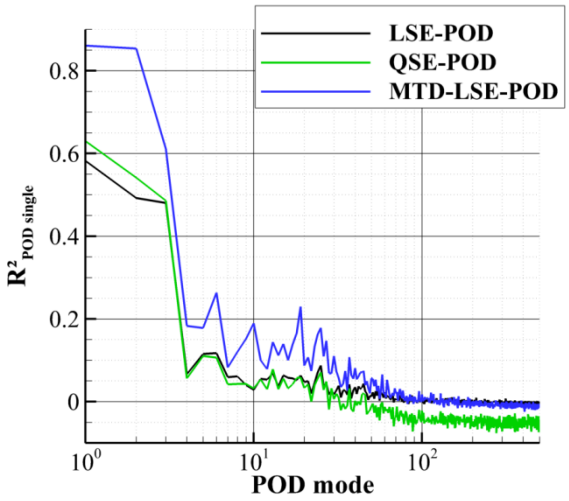


Fig. III.71: Determination coefficient calculated for each POD mode separately on the validation set using LSE, QSE and MTD-LSE-POD with 100 delays taken every 10^{-2} $T_{shedding}$ (Cube: low speed PIV, vertical plane)

- *Horizontal plane:*

The evolution of R^2 (calculated on the validation set) as function of the number of delays used for several constant time steps between consecutive delays is plotted in Fig. III.72. The same situations as for the vertical plane are tested. Once more these values are obtained using Tikhonov regularization and the regularization parameter is determined by cross-validation. The trends are perfectly consistent with those observed for the estimation of the vertical plane. The optimal time window still ranges from 0 to 0.75-1 $T_{shedding}$ depending on the time step between consecutive delays. The highest R^2 is 29.7% (13% higher than when using LSE). As in the previous test cases, MTD-LSE still holds the best prediction accuracy compared to single-time LSE and QSE.

The prediction by MTD-LSE of the vertical velocity of the same instantaneous field as in Fig. III.54 is plotted in Fig. III.73. Compared to the LSE and QSE predictions, the MTD-LSE prediction seems qualitatively better. Levels are higher than for the LSE predictions but not than the QSE predictions. The positive fluctuation structure spanning from (x/h = 4, y/h = -0.2) to (x/h = 5, y/h = 0.6) is better predicted by the MTD-LSE than by the other techniques.

The normalized mean square error is displayed in Fig. III.74 and Fig. III.75 for the three velocity components. These error maps can be compared with those of Fig. III.56 (for the LSE) and Fig. H.6 in annex G (for the QSE). Similarly to the vertical plane prediction, the error is decreased in the entire field. Looking at the shape of areas of lowest error, they are stretched downstream. The error is slightly more decreased in the most downstream regions.

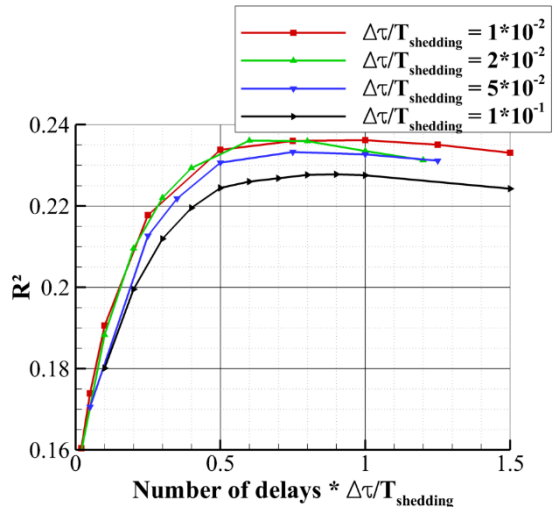


Fig. III.72: Determination coefficient as function of the number of delays for MTD-LSE estimations of the validation set (Low speed PIV cube experiment, horizontal plane)

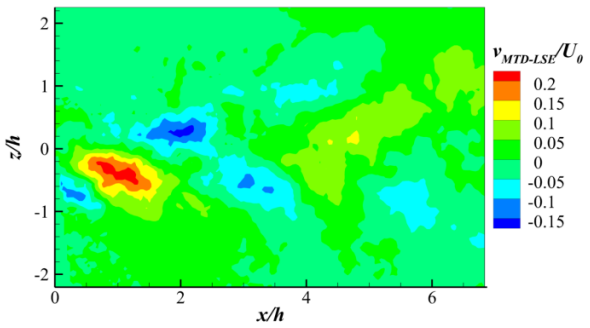
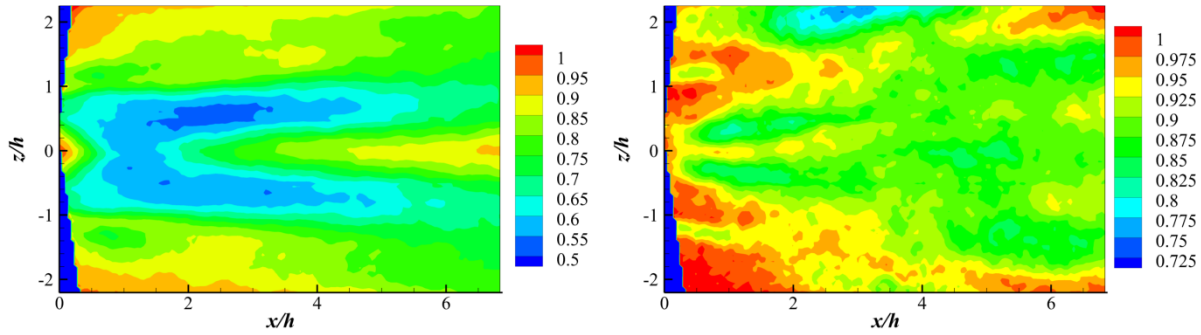


Fig. III.73: Instantaneous vertical velocity predicted by MTD-LSE (using 100 delays every 10^{-2} $T_{shedding}$)



a) b)

Fig. III.74: Normalized mean square error of the MTD-LSE prediction for the streamwise velocity (a) and the vertical velocity (b)

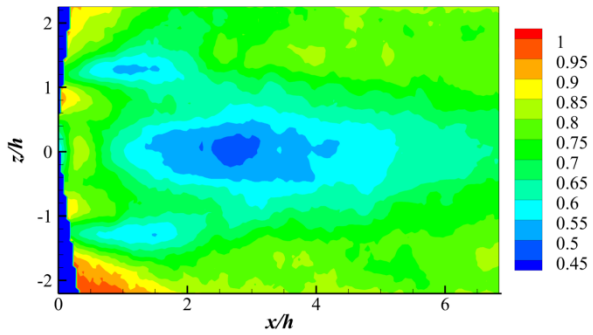


Fig. III.75: Normalized mean square error of the MTD-LSE prediction for the spanwise velocity

At last, the evolution of R^2 (from the validation set) for each POD mode from MTD-LSE-POD prediction is displayed in Fig. III.76. All POD modes up to rank 10 are far better predicted by MTD-LSE-POD than by LSE-POD or QSE-POD. Especially, the first three POD modes prediction are good with R^2 higher than 80%. As for the vertical plane, the prediction of POD modes of rank higher than 70 is not improved by the use of multi-time-delays. It is even deteriorated for some POD modes compared to the LSE-POD.

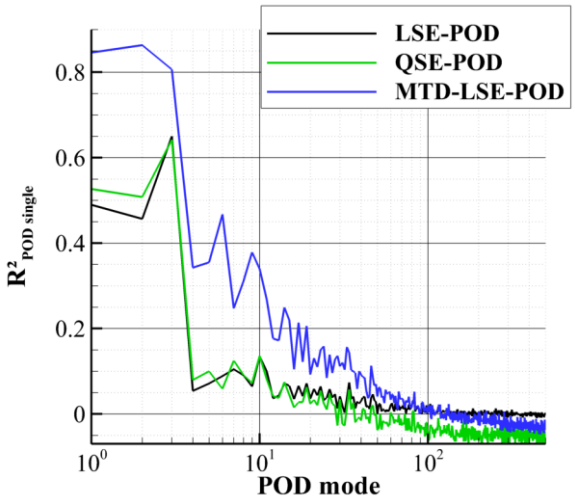


Fig. III.76: Determination coefficient calculated for each POD mode separately on the validation set using LSE, QSE and MTD-LSE-POD with 100 delays taken every 10^{-2} $T_{shedding}$ (Cube: low speed PIV, horizontal plane)

4.2. Application to the high speed PIV database

Using the high speed PIV database, the same conclusions as for the low speed PIV database are reached. Values of R^2 are summarized in Table III.15, Table III.16 (vertical plane) and Table III.17, Table III.18 (horizontal plane) for the LSE, LSE-POD, QSE and QSE-POD. Differences in R^2 values between the high speed PIV database results and the low speed ones come also from the differences that exist between the fields used in the two database (the size of the field is different and also the meshing). From these R^2 values, it is shown that the convergence of statistical moments necessary to the LSE and QSE is good enough. Indeed, the overfitting is not particularly stronger using the high speed PIV database than the low speed PIV one. That is also true when the upstream

velocity is 10 m.s^{-1} and 20 m.s^{-1} (see tables in annex H.4.2). However, for the horizontal plane, R^2 is higher on the validation set than on the training set in some cases. That is likely due to the limited size of the validation set. R^2 is probably not converged and one should not expect R^2 to be higher on the validation set than on the training set.

		LSE	LSE-POD (90%, 360 modes)	LSE-POD (70%, 71 modes)	LSE-POD (50%, 20 modes)	LSE-POD (33%, 6 modes)
R^2 from original fields	Training	14.7	14.7	14.6	14.1	13.1
	Validation	13.4	13.4	13.3	12.8	11.8
R^2 from POD projected fields	Training	-	16.3	20.8	27.9	38.5
	Validation	-	15	19.2	25.8	35.6
R^2_{POD}	Training	-	1.37	5.41	13.4	30.3
	Validation	-	1.16	5.12	12.6	28.2

Table III.15: Determination coefficients comparison between LSE and several LSE-POD (Cube: high speed PIV, vertical plane; values are expressed in %)

		QSE	QSE-POD (90%)	QSE-POD (70%)	QSE-POD (50%)	QSE-POD (33%)
R^2 from original fields	Training	18.5	18.4	17.9	16.8	15.1
	Validation	14.6	14.8	14.9	14.4	13.2
R^2 from POD projected fields	Training	-	20.4	25.6	33.3	44.4
	Validation	-	16.5	21.5	29.1	39.8
R^2_{POD}	Training	-	3.44	8.87	18.2	36
	Validation	-	0.28	5.56	14.7	31.8

Table III.16: Determination coefficients comparison between QSE and several QSE-POD (Cube: high speed PIV, vertical plane; values are expressed in %)

		LSE	LSE-POD (90%, 304 modes)	LSE-POD (70%, 54 modes)	LSE-POD (50%, 12 modes)	LSE-POD (36%, 3 modes)
R^2 from original fields	Training	21.8	21.7	21.6	20.7	19.4
	Validation	23.5	23.5	23.4	22.6	21.2
R^2 from POD projected fields	Training	-	24.2	30.7	41.2	52.9
	Validation	-	26.2	33.1	43.8	55.3
R^2_{POD}	Training	-	1.97	7.68	20.7	55.1
	Validation	-	1.91	8.11	23	58.6

Table III.17: Determination coefficients comparison between LSE and several LSE-POD (Cube: high speed PIV, horizontal plane; values are expressed in %)

		QSE	QSE-POD (90%)	QSE-POD (70%)	QSE-POD (50%)	QSE-POD (36%)
R^2 from original fields	Training	26.5	26.3	25.7	24	21.9
	Validation	24.2	24.3	24.4	23.6	22.1
R^2 from POD projected fields	Training	-	29.3	36.6	47.9	59.6
	Validation	-	27.2	34.6	45.8	57.4
R^2_{POD}	Training	-	4.38	12	26.9	60.8
	Validation	-	0.72	8.68	24.7	60.4

Table III.18: Determination coefficients comparison between QSE and several QSE-POD (Cube: high speed PIV, horizontal plane; values are expressed in %)

The evolution of R^2, calculated on the validation set, as function of the number of delays, for several MTD-LSE, is displayed in Fig. III.77 for the vertical plane and in Fig. III.78 for the horizontal plane. Evolutions of R^2 are perfectly similar to the ones obtained with the low speed PIV data. At last, R^2 calculated for each POD mode separately using LSE-POD, QSE-POD and MTD-LSE-POD are shown in Fig. III.79 for the vertical plane and in Fig. III.80 for the horizontal plane. Similar conclusions as when the low speed PIV data are used can be drawn.

In overall both databases lead to the same evaluation of SE methods accuracy and can therefore be used to characterize it.

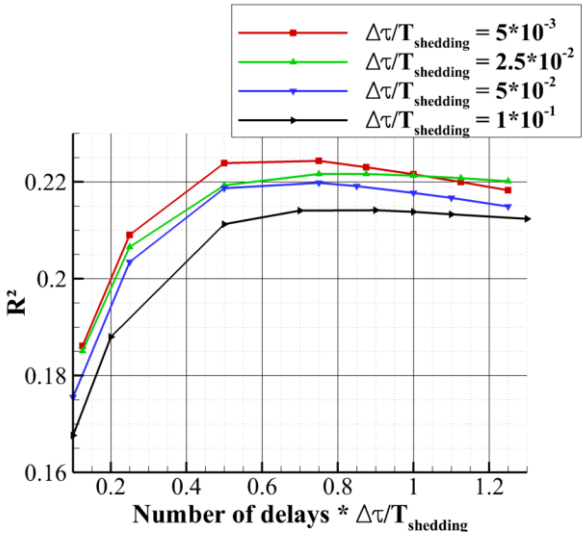


Fig. III.77: Determination coefficient as function of the number of delays for MTD-LSE estimations of the validation set (High speed PIV cube experiment, vertical plane)

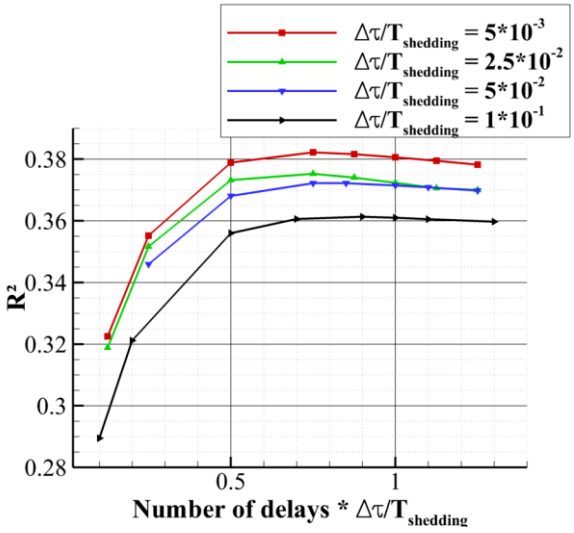


Fig. III.78: Determination coefficient as function of the number of delays for MTD-LSE estimations of the validation set (High speed PIV cube experiment, horizontal plane)

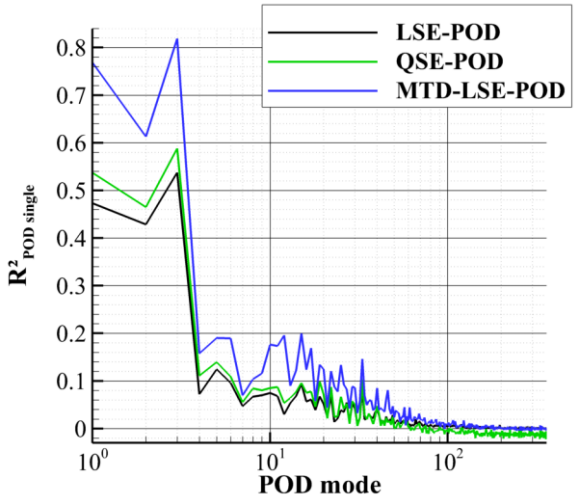


Fig. III.79: Determination coefficient calculated for each POD mode separately on the validation set using LSE, QSE and MTD-LSE-POD with 200 delays taken every 5×10^{-3} $T_{shedding}$ (Cube: high speed PIV, vertical plane)

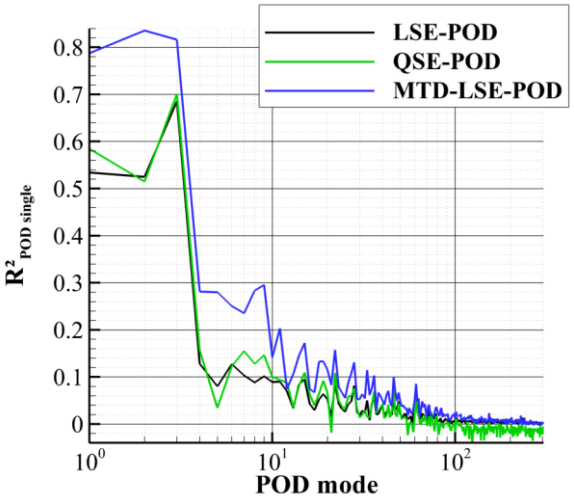


Fig. III.80: Determination coefficient calculated for each POD mode separately on the validation set using LSE, QSE and MTD-LSE-POD with 200 delays taken every 5×10^{-3} $T_{shedding}$ (Cube: high speed PIV, horizontal plane)

4.3. Summary

The comparison between the LSE, QSE and MTD-LSE leads to the exact same conclusions as in the BFS case. The best estimations are obtained with the MTD-LSE. The accuracy of SE prediction is slightly better than for the BFS case but remains quite lower than what was obtained for the blunt trailing edge case and OAT15A airfoil case. In addition, reconstruction and prediction quality is very different from one velocity component to another. For instance, the spanwise velocity, in the vertical plane, prediction R^2 is 23% while it is only 4.8% for the vertical velocity.

The study of the POD mode estimation shows that the first three POD modes (for the vertical plane or the horizontal plane) are estimated with relatively high accuracy (R^2 higher than 45%) whereas the estimation of POD modes of higher ranks is of poor quality (R^2 is mostly lower than 10%).

At last, it was observed that the high speed PIV campaign could also be used to evaluate the SE performances. But results on the validation sets have to be carefully discussed as the validation sets are not always large enough to ensure a good convergence of the error metric.

5. Chapter summary

Stochastic Estimation (SE) has been applied to the test cases available. For each of these cases, Linear SE, Quadratic SE, SE coupled with Proper-Orthogonal-Decomposition (POD) and also Multi-Time-Delays (MTD) SE have been used.

It was shown that looking at maps of the average correlation module between the conditional events and the data to estimate directly hints at where the SE will poorly perform, or at least allow to identify areas where SE will perform with less accuracy than in other areas. This information can therefore be retrieved without performing the estimation.

Reconstruction and prediction were found to be very accurate (whatever the SE methods used) in the blunt trailing edge flow case. Determination coefficient (R^2) is higher than 90% using only 4 pressure sensors. The good performances of SE to estimate a simple flow, with a dynamic that is composed only of the von Kármán vortex street, have been confirmed. Estimations are less accurate in the OAT15A case. R^2 is higher than 40% for all methods (using 3 pressure sensors). In particular, it has been observed that the vortex shedding developing in the wake of the airfoil was badly reconstructed and predicted by all SE methods. The shock oscillation is however well captured and the direct prediction of the shock prediction by LSE was found to be very accurate. For the estimation of the flow downstream the Backward Facing Step (BFS), it has been shown that the SE reconstruction and prediction accuracy was very poor. For this flow, R^2 remains lower than 10% using 17 pressure sensors and any SE methods. For the flow around a wall mounted cube test case, reconstruction and prediction are only slightly more accurate than in the BFS case with R^2 lower than 25% using 23 pressure sensors.

By comparing the accuracy of the several SE methods, it has been shown that QSE always holds more accurate reconstruction than LSE and could hold better prediction than LSE. However, for the BFS case, and in some cases for the wall mounted cube case, R^2 was smaller using QSE than LSE. That is the consequence of an overfitting of the model, mainly due to the bad convergence of statistical moments necessary to train the QSE. The use of the Tikhonov regularization only slightly improves the quality of the QSE prediction in such cases.

For all cases, it was found that MTD-LSE leads to the most accurate reconstruction and prediction between all tested methods. Also it is important to notice that the use of MTD-LSE corresponds to an increase in the number of conditional events. As such, in the OAT15A airfoil case using 3 pressure sensors, the LSE uses only 3 conditional events, the QSE can be considered to be using 6 conditional events, and the best MTD-LSE situations uses 210 conditional events.

Using MTD-LSE, R^2 reaches almost 100% in the blunt trailing edge case, 53% in the OAT15A case (with 3 pressure sensors), but only 9.3% in the BFS case. In addition, the good performances of the MTD-LSE require using Tikhonov regularization (with cross-validation for the choice of the regularization parameter). It was shown that the number of delays and the constant time step between them play an important role on the estimation accuracy. As such, delays have to be chosen inside a time window that usually goes in the past up to the main phenomenon period. Increasing the number of delays taken in the time window (thus decreasing the time step between them) increases R^2 on the validation set, but only if Tikhonov regularization is applied. Without regularization, while R^2 calculated on the training set keeps increasing when more and more delays are used, R^2 on the validation set goes by a maximum lower than when regularization is used (or decreases directly). At last, if the improvements are substantial in the blunt trailing edge case, OAT15A case and wall mounted cube case, they are very limited in the BFS case.

The use of LSE-POD or QSE-POD showed that these methods do not improve in themselves the reconstruction or prediction accuracy of the entire initial flow field. Nevertheless, it was possible to slightly improve the QSE prediction using QSE-POD. Indeed, it was observed that some POD modes suffer more from overfitting than others. Trying to predict them actually deteriorates the overall estimation. In these conditions, the possibility to disregard them, using QSE-POD, leads to the prediction improvement. The study of these methods also allowed highlighting than some POD modes are greatly better estimated than others. The interpretation of these observations will be addressed in the next chapter.

At last, the impact of the sensor locations was clearly identified in the trailing edge and the OAT15A test case. This subject is therefore addressed in the Chapter V.

Chapter IV. Characterization of Stochastic Estimation methods

The aim of this chapter is to characterize SE methods in order to better understand the differences that rose between the estimation of the several test cases in the previous chapter. In addition, this characterization may serve to highlight the limitations of SE methods and in particular to determine in which conditions they will perform with high accuracy.

First, the impact of Gaussian noise on the data is quickly evaluated and its impact on the evaluation of the SE method performance is demonstrated. Then the comparison, between the estimations of the flow around the OAT15A profile from the URANS and ZDES simulations, is used to point out some characteristics of SE in relation to the flow turbulence.

Following the observations made, the estimation of the turbulent spatial integral length scales is studied. The temporal filtering effect of SE is also examined, as well as the spatial filtering effect in term of Gaussian spatial filter.

The impact of increasing the Reynolds number on the estimation quality of same flow geometry is then analyzed, as well as the impact of training the SE at some upstream velocity conditions to reconstruct the flow at another upstream velocity condition.

At last, following the work of Ruiz et al., the use of Spatio-Temporal POD to determine a phase for the velocity field is investigated using the wall mounted cube case. This phase is used to propose a modification of the LSE.

1. Impact of Gaussian noise

Data from simulation does not suffer from noise contamination as it is the case for experimental data. Therefore, they can be used to analyze the impact of the noise on the SE accuracy. In addition, it is ordinary to train SE methods using experimental data and, more important, our goal is to evaluate the capacity of SE to be used in real time on a true application for which the conditional events will necessarily be experimentally measured. Here, only Gaussian noise is considered. The presence of noise on the validation set only or on both training set and validation set is investigated. The presence of noise on the conditional events only or on both the conditional events and the data to estimate is considered. At last, the study is performed using the data from the OAT15A airfoil URANS simulation (the database is described in Chapter II.4.2.1).

First, the case where noise is added on both data from the training and validation sets is considered. Fig. IV.1 compares the evolution of the determination coefficient (of the validation set), as function of the Signal-to-Noise Ratio (SNR) for the several situations using LSE with the set of 3 pressure sensors presented in Chapter III.2.1. The green curve (circle symbol) displays the determination coefficient in the case where noise is added only to the velocity fields and the determination coefficient is calculated using the original fields without noise as reference. In this situation, the determination coefficient remains constant when the SNR decreases and is equal to the value obtained without noise on any data. This result indicates that the noise introduced in the fields used for the training is completely filtered. This observation is consistent with the method which is based on the correlations between the conditional events and the data to estimate. Therefore, since the Gaussian noise is uncorrelated, it is filtered by the LSE. The orange curve (diamond symbol) represents the evolution of the determination coefficient in the same situation than the green curve except that it is calculated using the fields containing noise as reference. Since the noise is filtered by the LSE, the prediction diverges more and more from the noisy fields as the SNR decreases. In this situation, the performance of the LSE can be considered as underestimated.

The blue curve (delta symbol) corresponds to the case where noise is added only to the sensors (training and validation data). This time the prediction is directly deteriorated and the lower the SNR is, the more deteriorated the prediction is. As the sensors are used directly to predict the velocity fields, the presence of noise on these data impacts the quality of the prediction. The black

curve (left triangle symbol) represents the case where noise is added to both sensors and fields, and the determination coefficient is calculated from the fields without noise. The black and blue curves superposed themselves perfectly which indicates once more that the noise contained by the fields is filtered and only the noise added to the sensors deteriorates the prediction.

At last, the red curve (gradient symbol) corresponds to the case where noise is added on the sensors and the fields and the determination coefficient is calculated using the fields with noise as reference. This situation is the closest to the experimental case, where the fields and sensors contain some noise. It is clear, from Fig. IV.1, that it is the worst situation. The prediction is deteriorated because the sensors used contain noise and moreover the abilities of the LSE are underestimated as the determination coefficient is calculated from the fields that are polluted with noise.

Fig. IV.2 illustrates the same study run for the QSE. The exact same observations can be made as for the LSE.

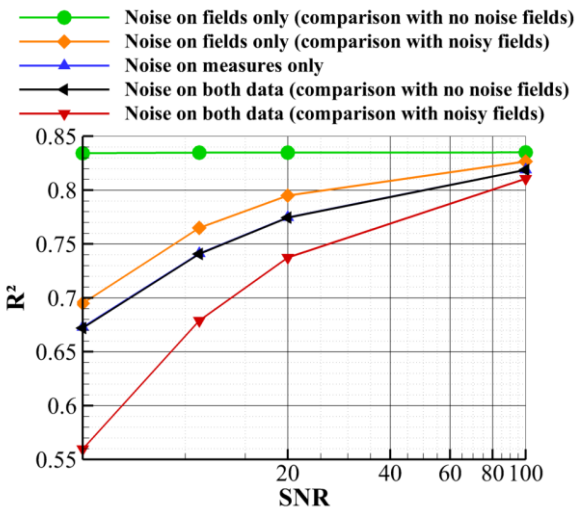


Fig. IV.1: Determination coefficient evolution as function of the SNR (LSE prediction with 3 pressure sensors, noise is added on both training and validation data)

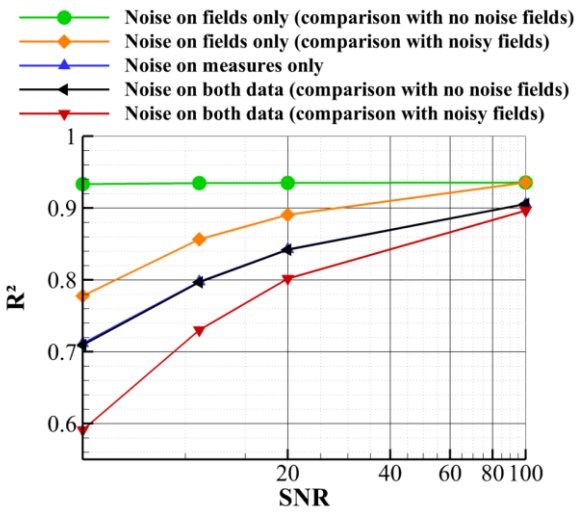


Fig. IV.2: Determination coefficient evolution as function of the SNR (QSE prediction with 3 pressure sensors, noise is added on both training and validation data)

Now the case where only data out of the training set contain noise is investigated. This situation could be encountered if the SE were trained on numerical data and then used on a real application using experimental measurements. In this situation, only noise contained by the sensors deteriorates the estimation. Indeed, it was shown that the noise contained by the field used to calculate R^2 leads to underestimate R^2 but did not lead to a true deterioration of the estimation. Therefore, only the addition of noise of the sensors is considered.

The results are shown in Fig. IV.3 where the evolution of R^2 is plotted as function of the SNR of the noise added to the sensors in the validation set (TR stands for Tikhonov regularization). Results from two sets of pressure sensors are compared (they are defined in Chapter III.2.1, one is composed of 3 sensors, the other of 5 sensors). For both sets, R^2 increases with the SNR, thus noise on the pressure measurements clearly deteriorates the prediction quality. In addition, the deterioration is more important in the case where only the validation set pressure measurements contain noise than in the case where both training and validation set measurements contain noise. That is seen by comparing the red curve in Fig. IV.1 with the green curve in Fig. IV.3. Also, the deterioration is extremely strong for the set of 5 pressure sensors. This is related to the collinearities between the sensors and thus to the conditioning of the covariance matrix E in equation (II.7). Indeed, the VIF obtained for the set of 3 sensors is 10.9 and for the set of 5 sensors 82. And the conditioning of

E is 265 with 3 sensors and 533 with 5 sensors. When the conditioning increases, the model becomes more sensible to perturbations in the measurements (input). Typically, the noise that the measurements contain can be amplified. It is what is observed here. As explained in Chapter II.1.4, in such situations, regularization can decrease the error outside the training set and thus improve the prediction accuracy of the model. This was quickly investigated here. Using Tikhonov regularization and choosing the regularization parameter through cross-validation, R^2 could be increased up do 68% instead of -104% in the case using 5 pressure sensors with SNR equals to 5. Results are plotted in Fig. IV.4. Using Tikhonov regularization, it can be seen that it is possible to almost match the performance obtained when noise measurements were used for the training.

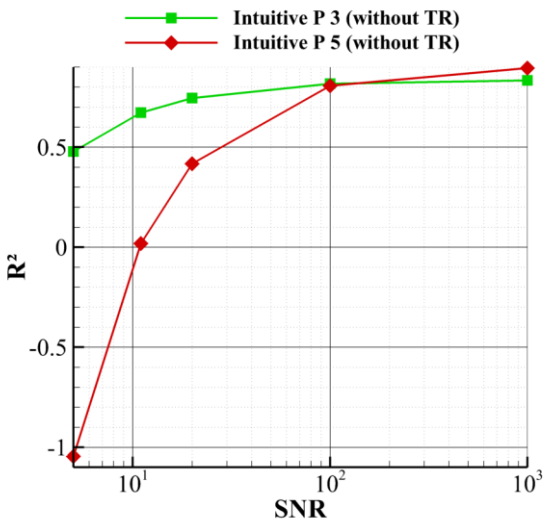


Fig. IV.3: Determination coefficient evolution as function of the SNR for LSE predictions using two sets of pressure sensors (noise is added on the pressure sensors on the validation set only)

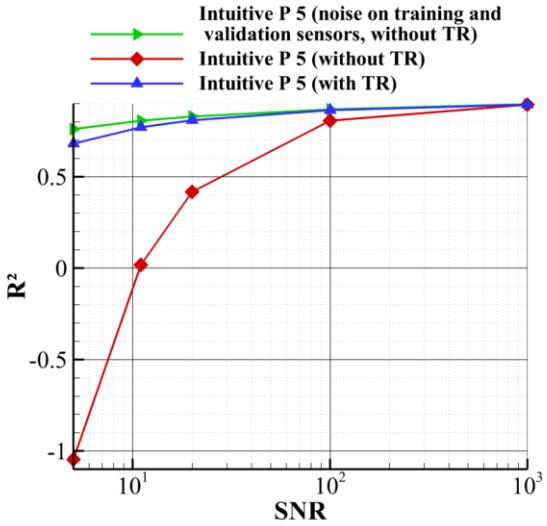


Fig. IV.4: Determination coefficient evolution as function of the SNR for LSE predictions using the set of 5 pressure sensors, with or without using Tikhonov regularization. Noise is added on the pressure sensor on both training and validation sets (green curve) or on validation set only (blue and red curves)

This study of Gaussian noise demonstrates the criticality of the noise that can be contained by the conditional events. Indeed, such noise directly deteriorates the prediction accuracy of the SE. In addition, careful attention is required when training the SE from numerical data only. Indeed, one will overestimate the prediction accuracy of the method (since it will be evaluated using measurements that do not contain noise). But it is also possible that the SE model becomes very sensible to noise and thus badly performs using experimental measurements. At last, it is interesting to notice that, if only experimental data are used, then the performance of the SE will be underestimated since the reference data contain noise that is filtered.

2. Comparison of estimations by SE of URANS and ZDES simulations of the OAT15A airfoil in transonic conditions

In this section the estimation of velocity field simulated by URANS and ZDES methods are compared. In the URANS method, all the turbulence is modeled, whereas in ZDES part of the turbulence is resolved. Therefore, the content of the ZDES simulation is closer to the reality especially in regards to the smallest scales of the flow. This comparison should highlight the impact of the turbulent content of the flow on the estimation quality.

2.1. Linear Stochastic Estimation (LSE)

As in Chapter III.2, static pressure data extracted from the simulation are used as conditional events for the estimation of the URANS simulated velocity field. The two sets of pressure sensors used are identical to those utilized with the ZDES results (see Chapter III.2.1).

Fig. IV.5 shows an example of a validation set streamwise fluctuating velocity field estimation. The top left frame (a) is the original velocity field, the top right (b) the prediction using 3 pressure sensors and the bottom one (c) the prediction using 5 pressure sensors. The first thing to notice is that there are no clear vortices in the profile wake, contrary to the ZDES case. That comes from the fact that the turbulence is completely modeled when using URANS equations. The vertical root-mean-square velocity is displayed in Fig. IV.6 and can be compared with Fig. III.19. The levels are clearly smaller in the URANS case than in the ZDES case. The RMS is actually decreased mainly in the wake showing that a large part of the turbulence is not simulated in this region, using URANS equations. The second thing to notice is that estimations from both sensor sets are qualitatively excellent, with some improvements when 5 sensors are used. In both prediction, the shock position is correctly recovered and the fluctuations levels are correct. It is clear from these figures that the LSE is able to correctly predict instantaneous streamwise velocity field of the URANS simulation from a very limited number of sensors (the same is also observed for the vertical velocity field).

Looking at the normalized mean square error plotted in Fig. IV.7 and Fig. IV.8, the LSE appears to be less accurate around the mean shock position and in the wake, in a similar fashion as when ZDES fields were estimated. However, the regions of high errors are far less stretched than in the ZDES case. It is especially true in the wake. Also, the addition of the 2 pressure sensors improves the prediction accuracy, as it was observed in the ZDES case.

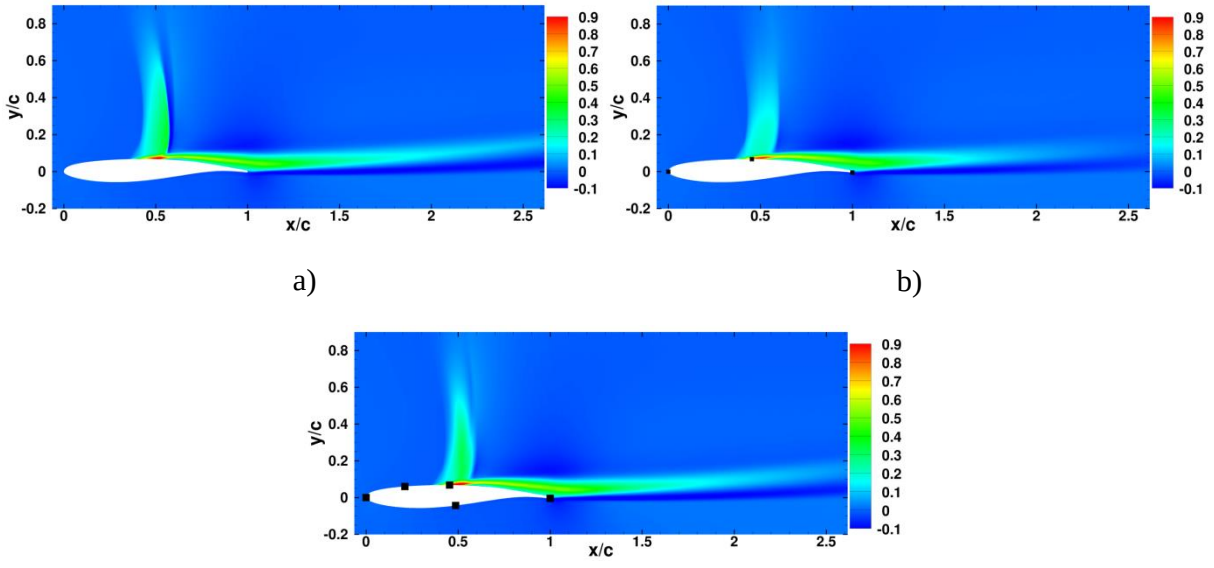


Fig. IV.5: Instantaneous fluctuating streamwise velocity (u/U_0). a) original field, b) LSE prediction with 3 pressure sensors, c) LSE prediction with 5 pressure sensors.

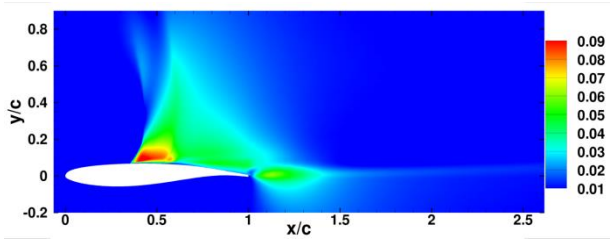
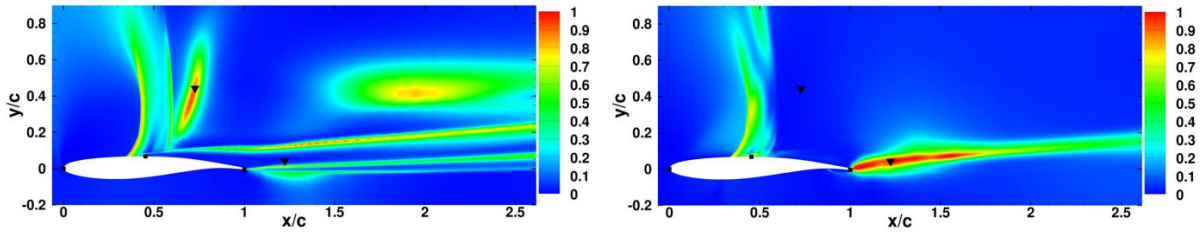
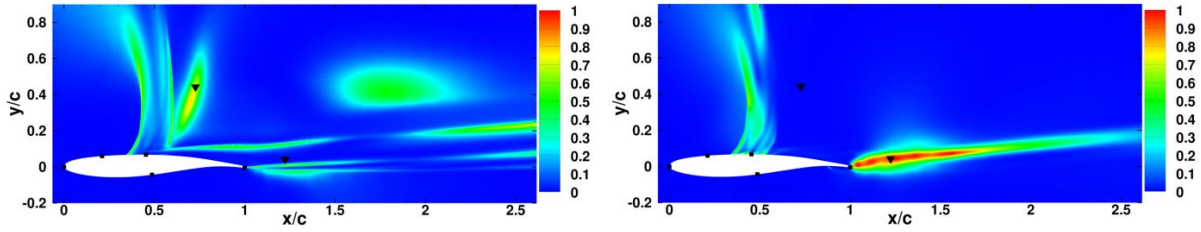


Fig. IV.6: Vertical root-mean-square velocity normalized by U_0



a) b)

Fig. IV.7: Normalized mean square error of the LSE prediction of the URANS velocity field using 3 pressure sensors. a) streamwise velocity component, b) vertical velocity component

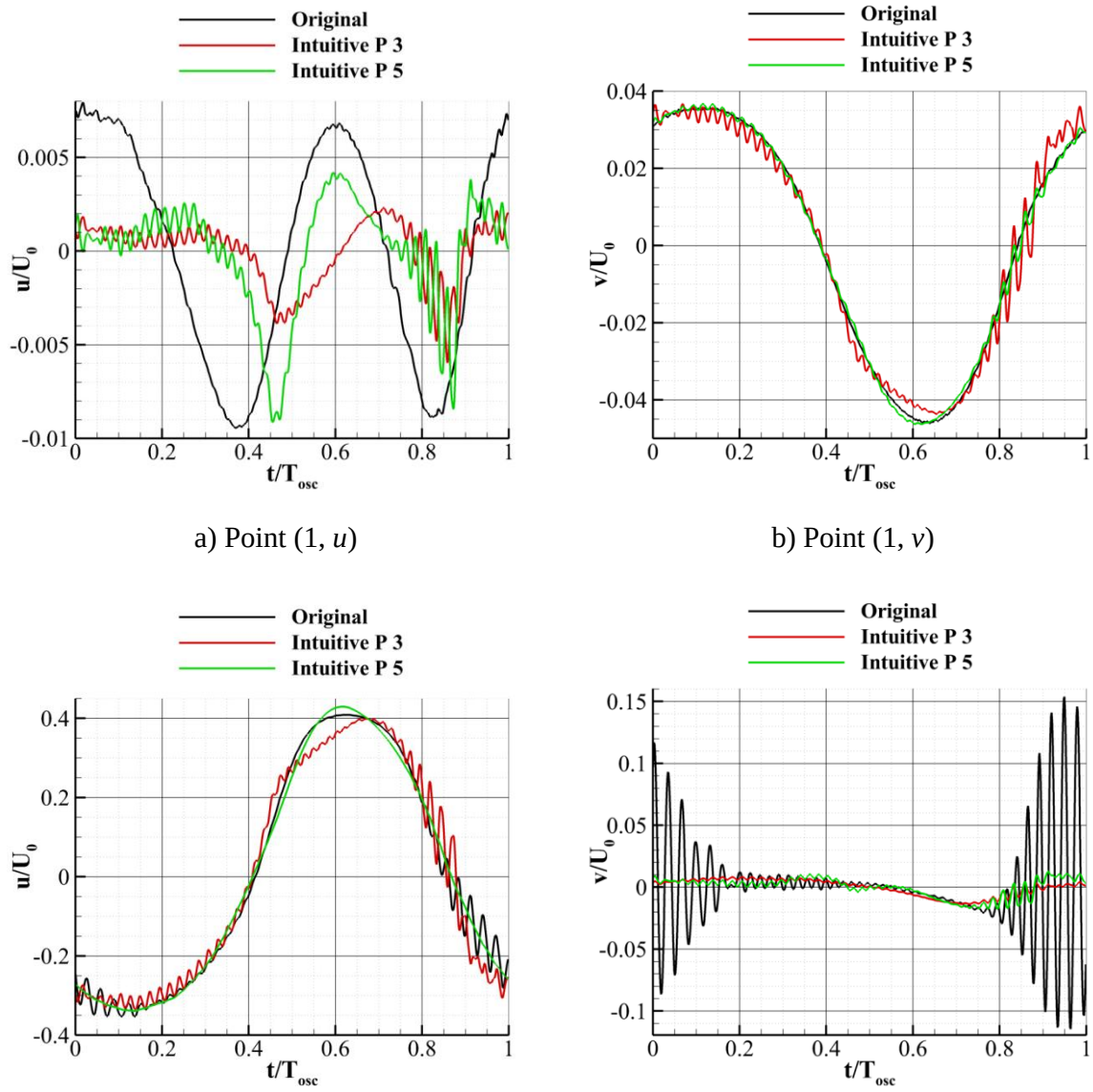


a) b)

Fig. IV.8: Normalized mean square error of the LSE prediction of the URANS velocity field using 5 pressure sensors. a) streamwise velocity component, b) vertical velocity component

The good estimation of the velocity fields from the URANS simulation is confirmed by the values of the determination coefficient obtained with both sets of sensors. 83.5% with 3 sensors and 90.4% with 5 sensors. These values are much higher than those obtained for the estimation of the ZDES velocity fields (43.6% and 53.2%). The estimation of the URANS velocity fields is more accurate than the estimation of ZDES velocity fields.

In order to better understand this observation and why some areas of the velocity field are estimated with less accuracy, a study of the temporal prediction of the velocity field is conducted. Signals of both *u* and *v* from the URANS simulation are displayed in Fig. IV.9. The positions of the two points, where *u* and *v* are extracted, are shown in Fig. IV.7 and Fig. IV.8 and correspond to the black triangles. The first point is located near the shock, in a region where *u* is not well predicted but the prediction of *v* is correct. The second point is located in the wake. There, *u* is well predicted but not *v*. That is clearly seen in the figures. Predicted signals at point (1, *u*) and point (2, *v*) do not match the original signals at all. For point (1, *u*), predictions present fast oscillations that are absent in the original signal. On the opposite, the fast oscillatory behavior of *v* at point 2 is absolutely not predicted by either set of pressure sensors. For *v* at point 1 and *u* at point 2, the prediction is improved by using 5 sensors. The main effect of the addition of 2 sensors on the prediction, for these cases, is to eliminate the high frequency content of the signal.



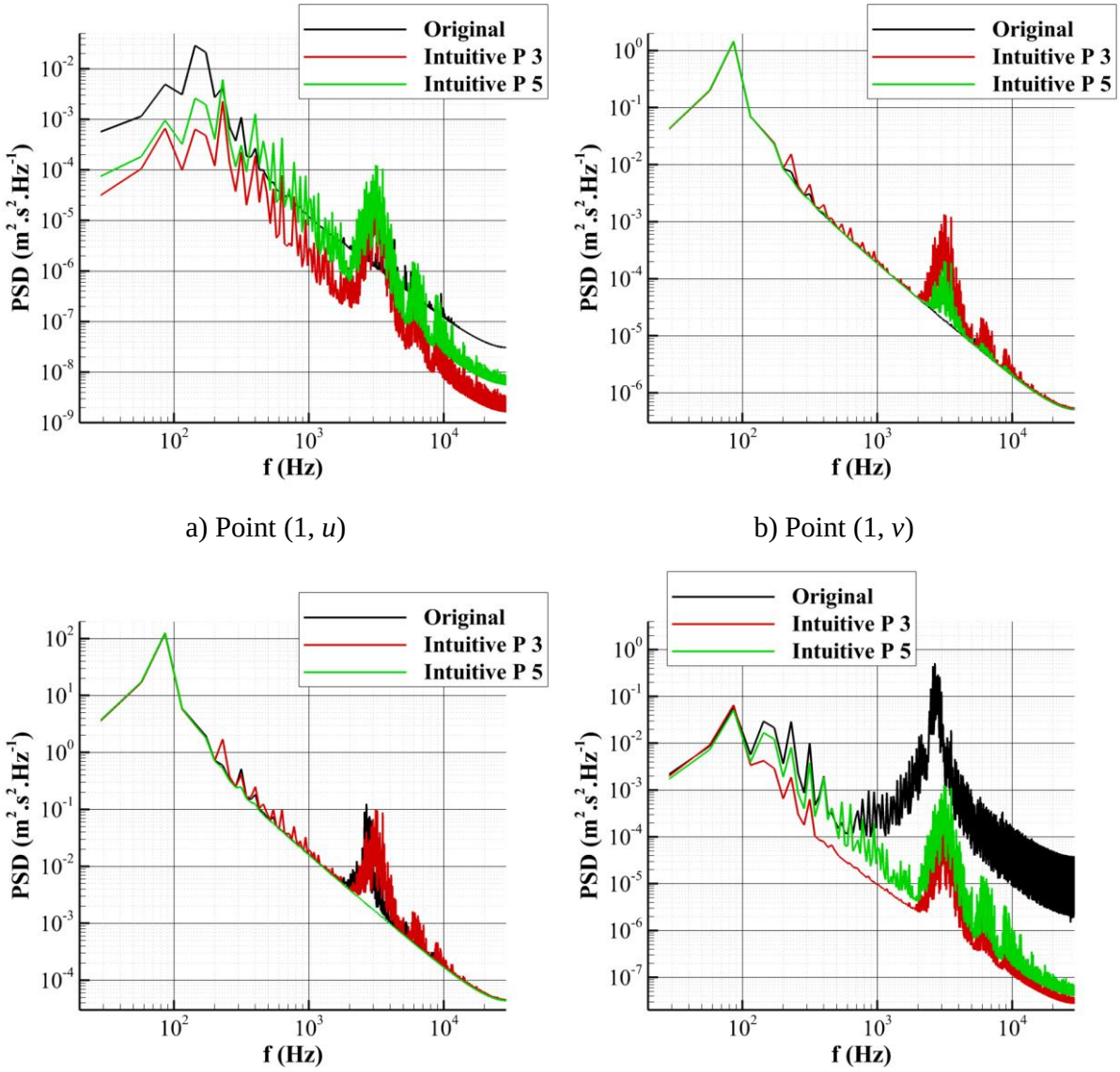
a) Point (1, *u*) b) Point (1, *v*)

c) Point (2, *u*) d) Point (2, *v*)

Fig. IV.9: Streamwise and vertical fluctuating velocity signals comparison between original data and LSE predictions (URANS case)

From these figures, it is obvious that the data best predicted present one main oscillation with a period equal to the shock oscillation period. On the other hand, when the data is badly predicted, it presents higher frequencies. That is confirmed by looking at the Power Spectral Density (PSD) of the signals in Fig. IV.10. The original signals of *u* at point 1 and of *v* at point 2 have clearly more energy in high frequencies than the two other signals. For point (1, *u*) the main frequency peak is at the first harmonic of the shock oscillation frequency. For point (2, *v*) a peak is observed around 3 kHz which corresponds to the vortex shedding frequency in the wake of the airfoil. For point (1, *u*) even the low frequencies are not correctly predicted, whatever set of sensors is used. In addition, the prediction has a high frequency content (about 3 kHz) that is not in the original signal. For point (2, *v*), the low frequency content (in particular at the oscillation frequency) seems correctly predicted and the use of 5 pressure sensors improves the prediction. The high frequency content (from 1 kHz) is, on the contrary, not predicted. This explains the bad performance of the LSE in this region. Fig. IV.10 shows that the LSE predicts with more accuracy the low frequency content than the high frequency one. It is important to notice here, that fitting of the estimated spectrum with the

original one, does not guarantee an accurate estimation. Indeed, no information is here given on the phase.



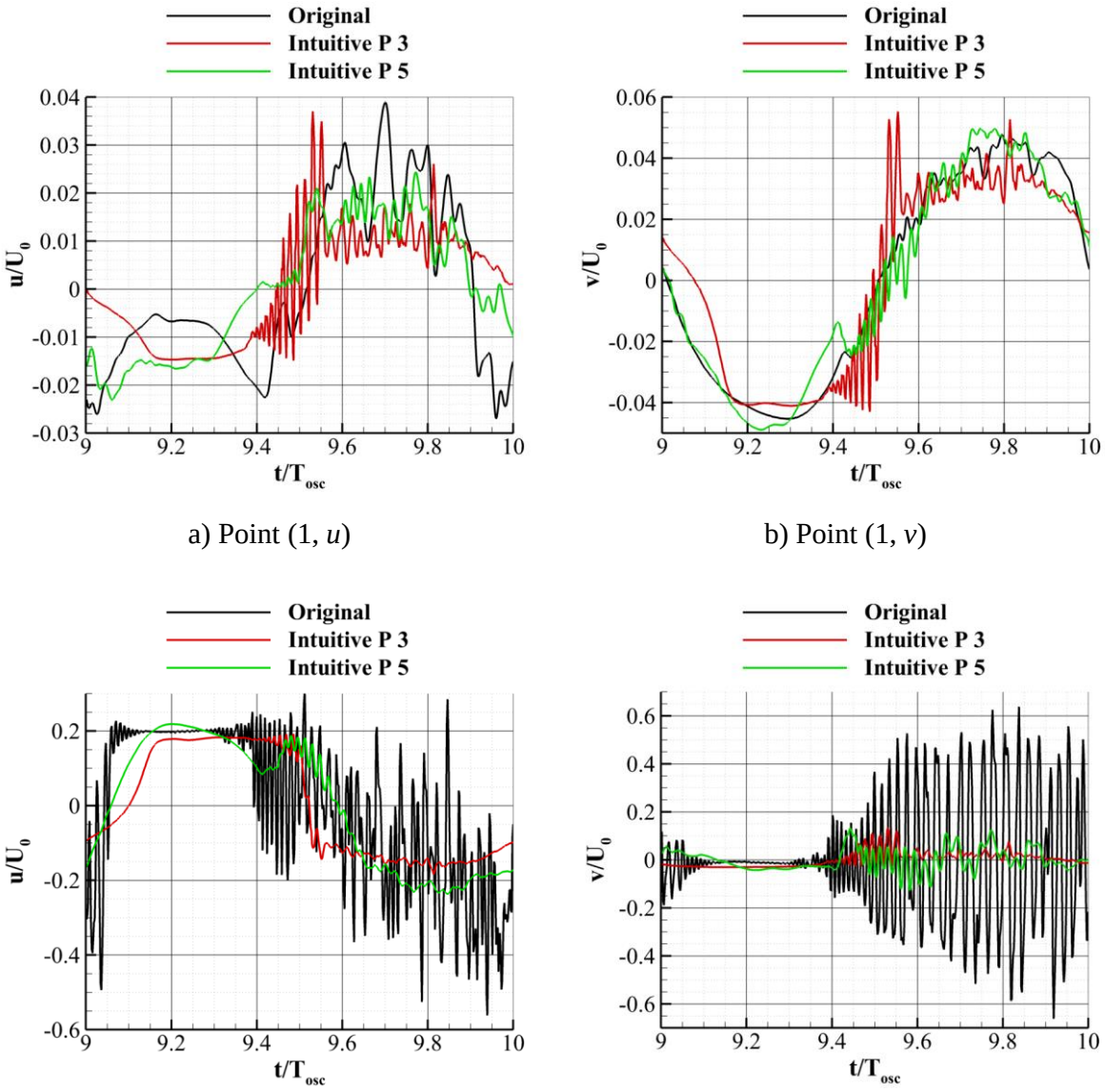
a) Point (1, *u*) b) Point (1, *v*)

c) Point (2, *u*) d) Point (2, *v*)

Fig. IV.10: Power Spectral Density comparison between the original and the LSE predictions using the sets of sensors “Intuitive P 3” and “Intuitive P 5” (URANS case)

The same signals are extracted from the validation set of the ZDES using LSE and the same sets of sensors. They are plotted in Fig. IV.11. Compared to the URANS signals, the ZDES ones are clearly more complex. If the signal at point 1 for *v* has still a clear main periodicity, it is not the case for point (1, *u*) and point (2, *u*). Predicted signals at point (1, *v*) globally follow the original signal and the main oscillation is recovered. However, they also contain fast oscillations that do not match the original signal. The same can approximately be said for point (1, *u*) and point (2, *u*), predicted signals seem to follow the slow evolution of the original signals but they absolutely do not match the original signals fast evolution. At last, the results for point (2, *v*) are similar to those obtained with URANS data. LSE predictions do not match the original data. Predictions however hold some fast oscillations which appear to be at the correct frequency but not particularly in phase with the original signal.

Power spectral densities (PSD) are shown in Fig. IV.12. From these spectra, it is clear that more energy is contained at high frequencies, especially at point 2. In addition, the low frequencies are better predicted than the high ones. For points (1, *u*), (1, *v*) and (2, *u*) the main peak at the shock oscillation frequency is well predicted. On the contrary, for points (1, *u*), (2, *u*) and (2, *v*) frequencies higher than 400 Hz are not correctly predicted for both sets of sensors. For these four signals, the prediction accuracy obtained in the ZDES case is lower than the accuracy obtained in the URANS case. The analysis of the spectral content of these signals suggests that the deterioration in accuracy comes from the fact that the high frequency content is not correctly predicted (in both URANS and ZDES case) and that more energy is contained in the high frequencies in the ZDES case than in the URANS case.



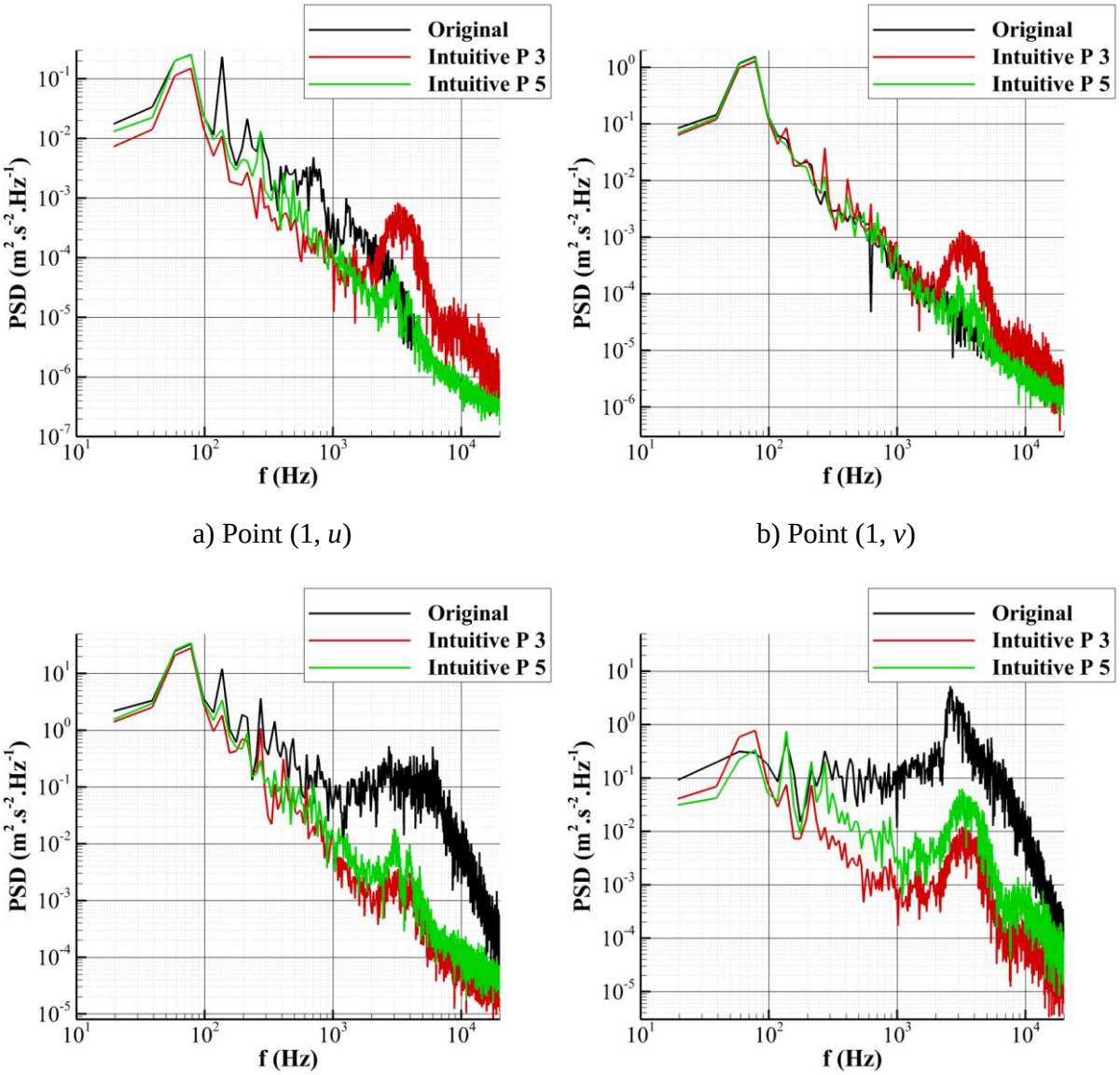
a) Point (1, *u*)

b) Point (1, *v*)

c) Point 2, *u*

d) Point 2, *v*

Fig. IV.11: Streamwise and vertical fluctuating velocity signals comparison between LSE predictions (ZDES case)



a) Point (1, *u*) b) Point (1, *v*)

c) Point (2, *u*) d) Point (2, *v*)

Fig. IV.12: Power Spectral Density comparison between the original and the LSE predictions using the sets of sensors “Intuitive P 3” and “Intuitive P 5” (ZDES case)

2.2. Modified Linear Stochastic Estimation (LSE-POD)

POD is directly computed using only training data from the URANS database. Fig. IV.13 shows the energy spectrum of the decomposition for the first hundred modes. This spectrum can be compared with the one from the ZDES database in Fig. III.22. The first POD mode accounts alone for more than 63% of the Turbulent Kinetic Energy (TKE) and the first four POD modes for almost 93%. In the ZDES case, the first POD mode contains 48% of TKE and the first four POD modes 79%. Thus, less POD modes are required to retain the same amount of TKE for the URANS database than for the ZDES database.

Determination coefficients of the first 10 POD modes of the URANS simulation (~ 98% TKE), obtained by LSE on the validation set, are plotted in Fig. IV.14 for both sets of pressure sensors. This figure can be compared with Fig. III.23 for the ZDES case. If the first two POD modes

are well predicted by both sensor sets (R^2 higher than 95%), from the third POD modes, the prediction drastically deteriorates (R^2 lower than 75%). However, the POD mode 6 is better predicted by both set of sensors. The use of "intuitive P 5" set greatly improves the prediction accuracy of POD modes 3 and in a smaller amount of POD modes 4, 5 and 6. Similarly to the ZDES case, only the first two POD modes are excellently predicted. Thus, the fact that the overall prediction of the velocity field is better for the URANS database than the ZDES database comes from the fact that the first two POD modes do not retain the same amount of TKE in both cases. In the URANS test case, the first two POD modes represent more than 84% of the TKE. In the ZDES case, these POD modes contain only 64% of the TKE.

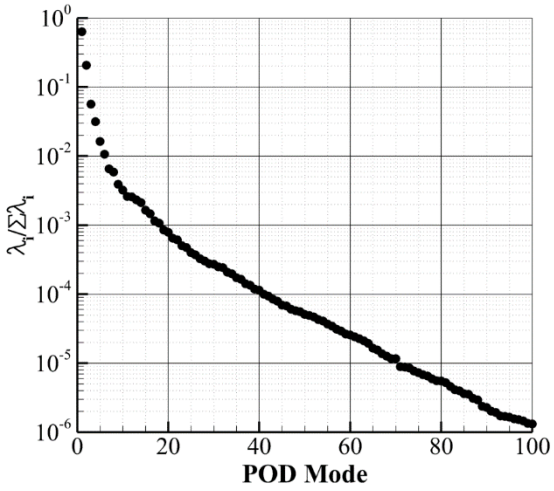


Fig. IV.13: Energy spectrum of the first hundred POD modes (OAT15A URANS)

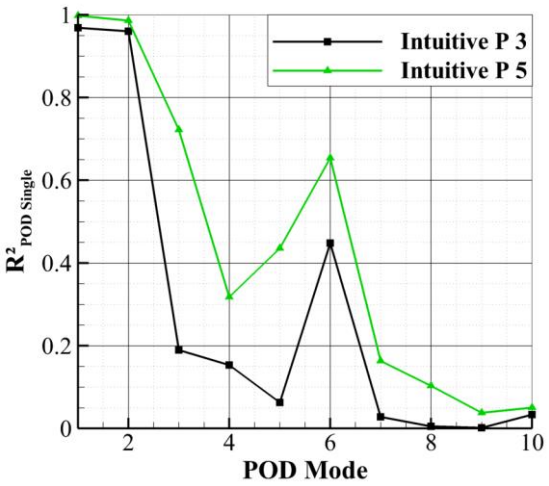
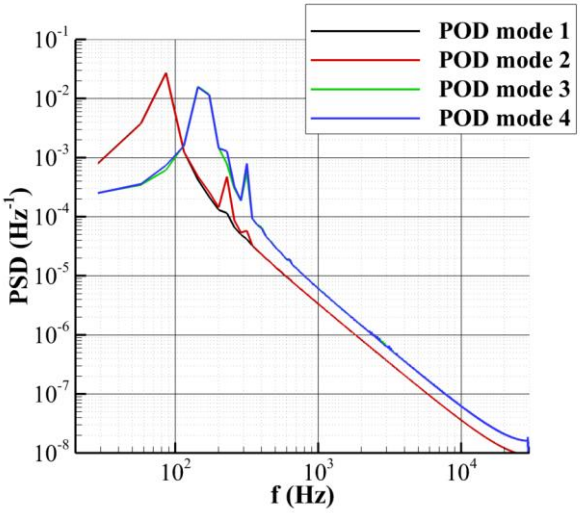
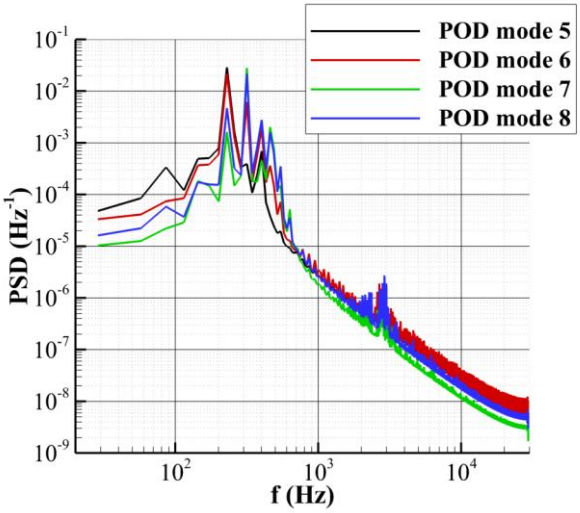


Fig. IV.14: Determination coefficient calculated for each POD mode separately on the validation set (OAT15A URANS)



a) POD modes 1 to 4



b) POD modes 5 to 8

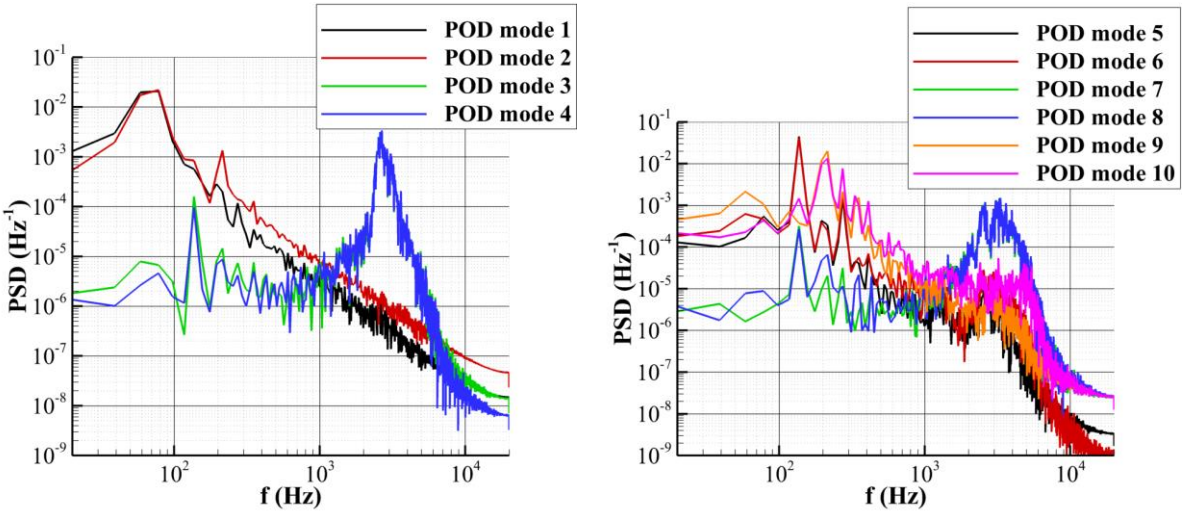
Fig. IV.15: Power Spectral Density of the POD modes 1 to 8 (OAT15A URANS)

The spectral content of the first 8 POD modes are displayed in Fig. IV.15 for the URANS case and in Fig. IV.16 for the ZDES case. In the URANS case, the first two POD modes correspond to the shock oscillation, with a peak at the shock oscillation frequency. Then higher POD modes

have more energy at higher frequencies. POD modes 3 and 4 corresponds to the first harmonic of the shock oscillation and the following modes to the superposition of higher harmonics. In this situation, the LSE seems to be able to only predict the shock oscillation (the first two POD modes). But since the shock oscillation accounts for a large part of the TKE, in the URANS case, the global prediction accuracy evaluated by R^2 is very good.

In the ZDES case, POD modes 1 and 2 still correspond to the shock oscillation, with a main peak at the shock oscillation frequency. POD modes 3 and 4 do not correspond to the first harmonic of the shock oscillation (as in the URANS case). They correspond to the vortex shedding in the wake of the profile, which frequency is 3 kHz. Concerning POD modes 5 to 8, POD modes 5 and 6 have more energy in the low frequencies than POD modes 7 and 8 (blue and green curves are superimposed for f > 1 kHz). It was already seen in Chapter III.2.1.2 that POD modes 1 and 2 were very well predicted. POD modes 5 and 6 were also slightly better predicted than POD modes 3, 4, 7 and 8. The fact that the POD modes 3, 4, 7 and 8 are poorly estimated confirms the inability of the LSE, with the chosen sensors, to estimate the high frequencies content of the flow. More precisely, the LSE is able to reconstruct and predict the buffet, but not the vortex shedding. Therefore, low frequency, coherent phenomena, such as the buffet, are more easily estimated by LSE, with a limited number of sensors, than others phenomena.

From the observations in the URANS and ZDES cases, it is possible to deduce the general trend that the higher in frequencies the energy content of one POD mode is, the more difficult it is to reconstruct (or predict) using SE. However, such deduction from the spectral analysis of the POD basis remains limited. Indeed, it is not possible to explain, from this information, why the POD mode 6 is better predicted than the others (in the URANS case) or why the POD mode 4 is better predicted than the third, with the "intuitive P 5" set in the ZDES case. These observations are addressed in Chapter V.1.2.1 where the impact of the sensor locations is investigated.



a) POD modes 1 to 4 b) POD modes 5 to 10

Fig. IV.16: Power Spectral Density of the POD modes 1 to 10 (OAT15A ZDES)

2.3. Quadratic Stochastic Estimation (QSE) and modified Quadratic Stochastic Estimation (QSE-POD)

As in the ZDES case, the reconstruction and prediction accuracy is improved by the QSE compared to the LSE in the URANS case. Determination coefficient, on the validation set, reaches 93.6% with the "Intuitive P 3" set of sensors and 97.9% with the "Intuitive P 5" set of sensors.

Determination coefficients of the first 10 POD modes, obtained by QSE on the validation set, are plotted in Fig. IV.17 for both sets of pressure sensors. LSE values are also plotted and the figure can be compared with the one of the ZDES case in Fig. III.23. QSE seems to improve the prediction of the first POD modes, especially when the "intuitive P 5" set is used. However, the prediction accuracy decreases for POD modes of rank higher than 7. In the ZDES case, the prediction of the first 10 POD modes is also improved by using QSE instead of LSE (see Fig. III.23). But, the improvement is much important for POD modes 5, 6, 9 and 10. It was previously shown that these POD modes contain more energy in the low frequencies of the flow than POD modes 3, 4, 7 and 8 (see Fig. IV.16). Thus, the QSE is also unable to predict the vortex shedding in the ZDES case and similarly to LSE, low frequencies are better estimated than high frequencies. These conclusions are not in opposition with the URANS results. Indeed, in the URANS case, up to POD modes 8, POD modes contain mainly low frequencies (under 1 kHz).

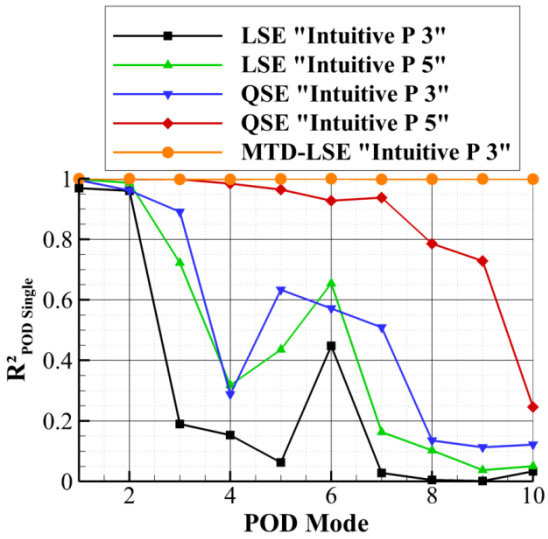


Fig. IV.17: Determination coefficient calculated for each POD mode separately on the validation set (OAT15A URANS) (in MTD-LSE 150 delays every 6.67×10^{-3} T_{osc} in the past are used with Tikhonov regularization)

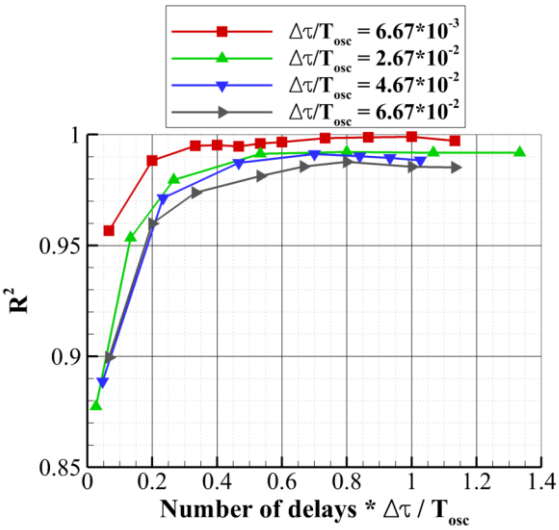


Fig. IV.18: Validation set determination coefficients as function of the number of delays for MTD-LSE (OAT15A URANS case)

2.4. Multi-Time-Delay Linear Stochastic Estimation (MTD-LSE) and MTD-LSE-POD

Using only "intuitive P 3" set of sensors, several constant time steps between delays are considered, as well as different numbers of delays. Fig. IV.18 shows the evolution of R^2 (from the validation set) with the number of delays for several constant time steps between delays. The values are obtained using Tikhonov regularization. Without applying Tikhonov regularization, as in the ZDES case, validation set R^2 mostly decreases when more delays are used and the prediction accuracy is never as good as when Tikhonov regularization is employed. On the contrary, without regularization, the reconstruction (training set estimation) accuracy improves when more delays are used. Once more an optimal time window is observed and spans from 0 to 1 T_{osc} approximately.

R^2 reaches a maximum of 99.9% when 150 delays taken every 6.67×10^{-3} T_{osc} in the past are used. As in the ZDES case, MTD-LSE is able to greatly improve the estimation of the velocity field. Here, MTD-LSE allows a perfect prediction of the velocity field even when only 3 pressure sensors are used. R^2 of the prediction of the first 10 POD modes using MTD-LSE with the same delays is

plotted in Fig. IV.17. The first 10 POD modes are all perfectly predicted. In URANS case, the high frequency content of the flow is well predicted using MTD-LSE. This is also illustrated in Fig. IV.19 where the predicted signal using MTD-LSE at point (2, *v*) is displayed.

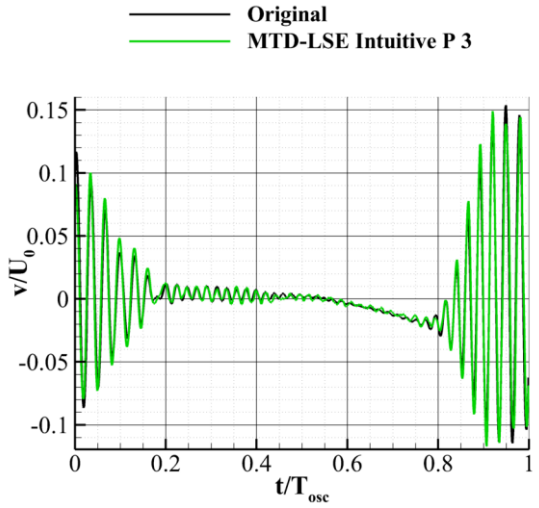
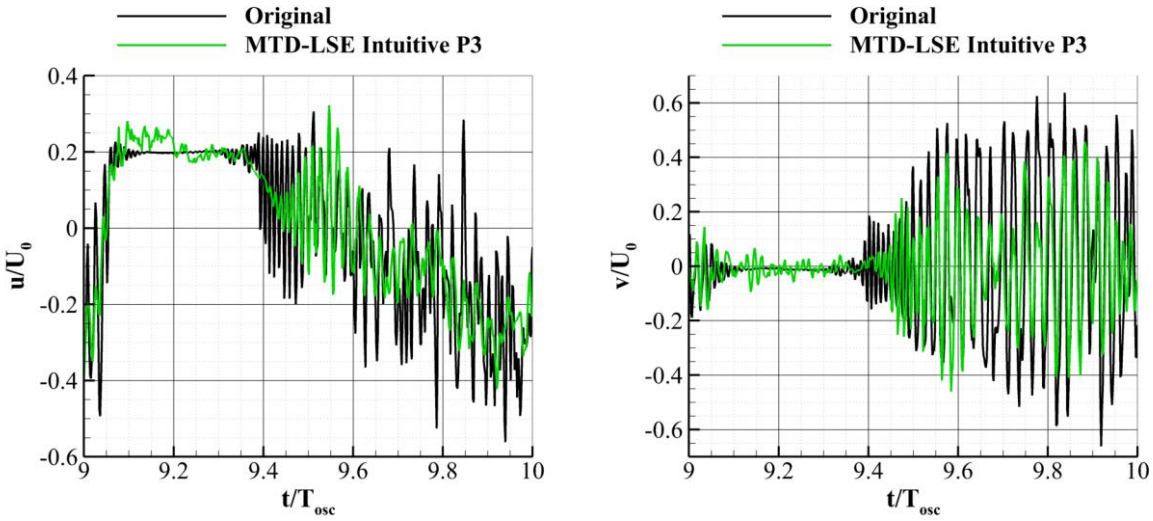


Fig. IV.19: Vertical fluctuating velocity signals comparison between original data and MTD-LSE predictions using 150 delays every 6.67×10^{-3} T_{osc} (URANS case, point 2)



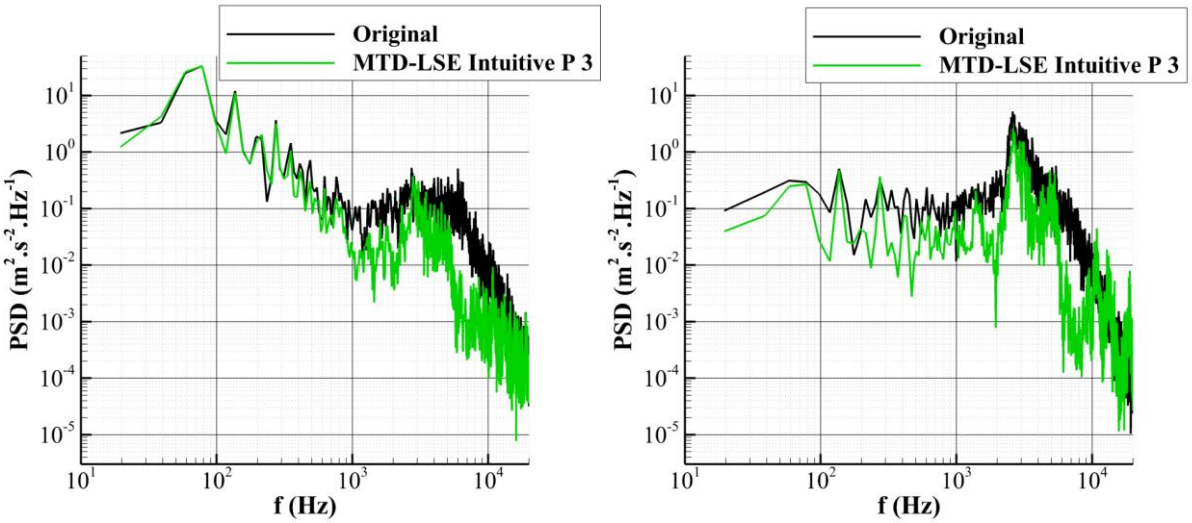
a) Point (2, *u*) b) Point (2, *v*)

Fig. IV.20: Streamwise and vertical fluctuating velocity signals comparison between MTD-LSE prediction (150 delays every 6.67×10^{-3} T_{osc}) and original (ZDES case)

It was shown in the previous chapter that MTD-LSE improved the prediction in the ZDES case. However, even in the best situations tested, the determination coefficient remains lower than in the URANS case (about 78%). Looking at Fig. III.26, MTD-LSE is shown to be less accurate for POD modes that contain more energy in the high frequencies (3, 4, 7 and 8). Therefore, in the ZDES case, even with MTD-LSE it is not possible to perfectly predict the high frequency content of the velocity field. Predicted signals at point 2 are displayed in Fig. IV.20 and can be compared with Fig. IV.11. The amplitude of the predicted oscillations is clearly higher when MTD-LSE is used than when LSE is used. In addition, they appear to be mostly in phase with the original ones. Nevertheless, predicted and original signals are far from matching each other. Looking at the spectral

content in Fig. IV.21, it is clear that low frequencies and high frequencies are better estimated with the MTD-LSE than with the LSE. But spectra do not match in the high frequencies domain (f > 1 kHz) which is consistent with the lower values of R^2 obtained in the ZDES case compared to URANS case values.

The excellent results in the URANS case may come from the fact that the turbulence is completely modeled in this case. Thus, the original data are completely deterministic while, in the ZDES case, the velocity evolution in the wake possesses a more random behavior (part of the turbulence is directly resolved).



a) Point (2, u) b) Point (2, v)

Fig. IV.21: Power Spectral Density comparison between the original and the MTD-LSE predictions using 150 delays every 6.67×10^{-3} T_{osc} (ZDES case)

2.5. Summary

Estimations, using several SE methods, of the velocity field around the OAT15A airfoil in transonic conditions, from URANS and ZDES simulations, have been compared. If the estimation performances of the SE methods in the URANS case are excellent, they deteriorate in the ZDES case.

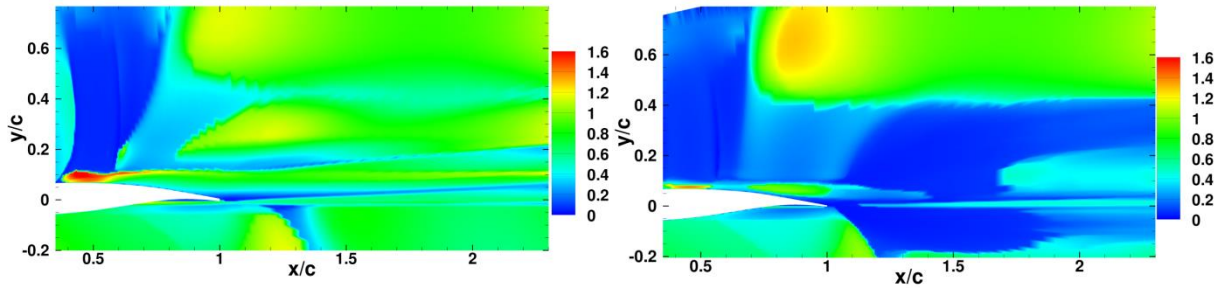
The spectral study demonstrated that SE methods have difficulty to correctly estimate the high frequency content of the flow. Especially, only in some cases of MTD-LSE (or MTD-LSE-POD), was the vortex shedding partially recovered. In the URANS case, the high frequency content represents a small part of the energy content of the flow. Indeed, the first four POD modes, that accounts for more than 93% of the Turbulent Kinetic Energy (TKE), presents one frequency peak around 80 Hz (for POD modes 1 and 2) and 160 Hz (for POD modes 3 and 4). These four POD modes are correctly reconstructed and predicted using LSE (or derived methods). But, in the ZDES case, the high frequencies represent a greater part of the flow TKE. The POD shows that, if the first two POD modes, accounting for 63% of the TKE, have a main frequency peak around 80 Hz, POD modes 3 and 4, which accounts for about 15% of the TKE, have a main frequency peak around 3 kHz, which is the vortex shedding frequency. The first two are well estimated but not the other two (even when using MTD-LSE-POD). These observations do not mean that, between two different flows, the one having more energy in the high frequencies will be estimated with the lowest accuracy using SE. But that for one flow, the low frequency content will be better estimated than

the high frequency content. The effect of SE on the spectral content will be addressed more precisely in Chapter IV.4.

Another aspect of turbulent flows concerns the turbulent spatial integral length scales which is linked to the two points correlation function [80] (formula are given in annex C). Thus, it can be of particular interest when using LSE (and its extensions) which is based on the correlations between the sensors and the flow field. Therefore, one could fairly expect that flow regions with the largest length scales are also the regions where the LSE performs the best.

Fig. IV.22 compares the turbulent spatial integral length scales of the streamwise velocity, calculated in the streamwise direction, of the URANS simulated flow field and of the ZDES one. The scales are particularly small in the area of displacement of the shock. It is due to the fact that the shock disrupts the correlation between points upstream and downstream the shock. In overall, comparing the two figures shows that the scales are shorter for the ZDES field than for the URANS one. In the wake region, the length scales are particularly shortened between the URANS case and ZDES case. It is due to the turbulence that is resolved in the last case and thus generates small scale turbulent structures in the wake. Those structures form the high frequency content of the vortex shedding.

This figure can also be compared with Fig. IV.8.a) and Fig. III.16.a) where the normalized mean square error of the streamwise velocity for the URANS case and ZDES case are displayed. Regions of short integral length scales match rather well the regions of high error. Especially the enlargement of the high error region in the wake, in the ZDES case compared to the URANS case, can be linked to the shortening of the spatial integral length scales in this area. Therefore, the LSE is unable to correctly estimate regions of the flow where the turbulent spatial length scales are the shortest and where the high frequency content is predominant. This effect is examined more closely in the next part.



a) URANS case

b) ZDES case

Fig. IV.22: Turbulent spatial integral length scales of the streamwise velocity in the streamwise direction (L_x/c)

3. Study of length scales conservation on the Backward Facing Step case

The Backward Facing Step (BFS) database of the low speed PIV experiments for an upstream velocity of 30 $m.s^{-1}$ is used to conduct the following study in the single-time case. When MTD is used, the high speed PIV database is employed.

3.1. Single-time Stochastic Estimation using wall pressure measurements

Fig. IV.23 and Fig. IV.24 show the turbulent spatial integral length scales for the streamwise velocity *u* and the vertical velocity *v* respectively (computed in both streamwise and vertical directions). The length scales are calculated as described in annex D (thus when calculated in the streamwise direction upstream and downstream length scales are taken into account, when calculated in the vertical direction, length scales calculated for descending *y* and for ascending *y* are taken into account). Since the integral length scales are computed on a finite domain, with a criterion for stopping the integration, they may not be perfectly converged in some points of the field. Indeed, the spatial domain may not always be large enough for the autocorrelation function to reach and remain close to 0 (in particular near the edges of the domain but the combination of Λ^+ and Λ^- should limit this problem). Also, according to Pope [96], the spatial domain should be at least six times larger than the maximum integral length value to accurately determine such integral length scale. However, we are interested in the comparison of original values with estimated ones, not in the integral length scale value in itself. Convergence issues are not expected to false the comparison.

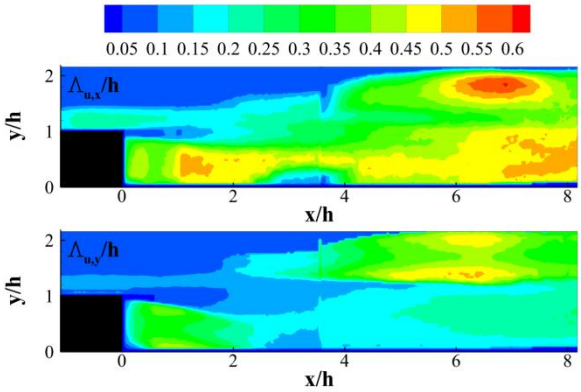


Fig. IV.23: Turbulent spatial integral length scale for the streamwise velocity *u*, computed in the streamwise direction (top) and vertical direction (bottom)

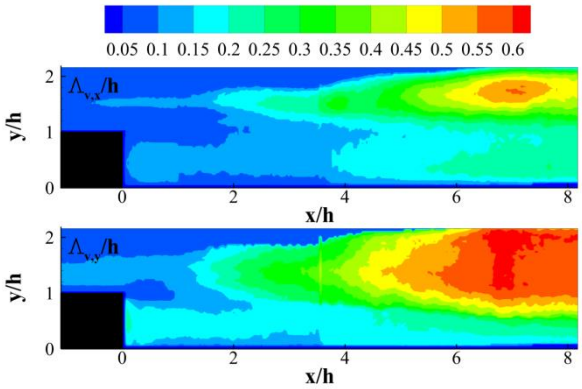


Fig. IV.24: Turbulent spatial integral length scale for the streamwise velocity *v*, computed in the streamwise direction (top) and vertical direction (bottom)

The discontinuities that can be observed around $x/h = 3.5$ (in Fig. IV.23 and Fig. IV.24) 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. For length scales calculated in the streamwise direction, the spatial domain seems large enough to evaluate the largest integral length scales. That is not the case for the scales calculated in the vertical direction, but the length scales calculated are consistent with the physics of the flow and in particular the vortex shedding phenomenon.

Estimated length scales (calculated for the validation set only) are plotted in Fig. IV.25 and Fig. IV.26. 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. IV.23 and Fig. IV.24 cannot be identified in the maps obtained from the estimated fields. The turbulent spatial integral length scales are not conserved by this LSE.

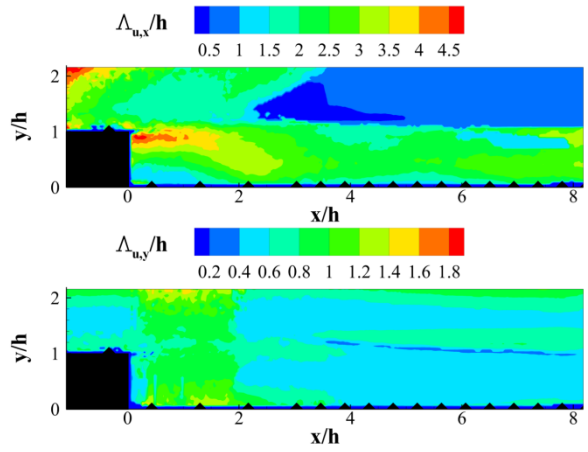


Fig. IV.25: Turbulent spatial integral length scale predicted by LSE, for the streamwise velocity *u* (top: computed in streamwise direction; bottom: computed in vertical direction)

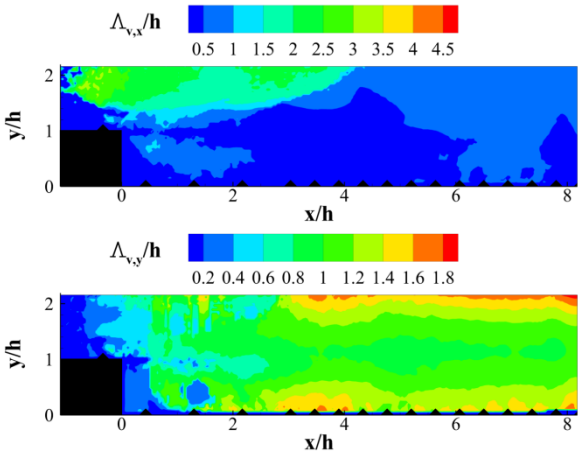


Fig. IV.26: Turbulent spatial integral length scale predicted by LSE, for the vertical velocity *v* (top: computed in streamwise direction; bottom: computed in vertical direction)

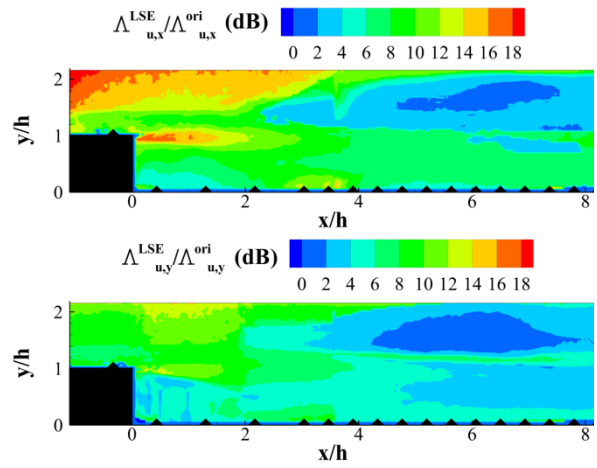


Fig. IV.27: LSE prediction to original turbulent spatial integral length scale for *u* ratio (top: calculated in streamwise direction; bottom: calculated in vertical direction)

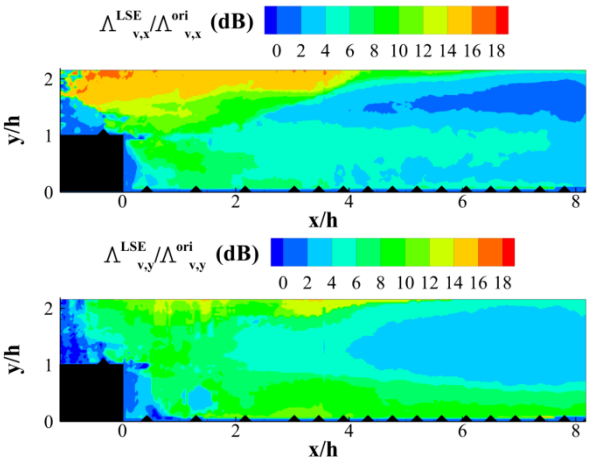


Fig. IV.28: LSE prediction to original turbulent spatial integral length scale for *v* ratio (top: calculated in streamwise direction; bottom: calculated in vertical direction)

In order to better compare the integral length scales obtained from the predicted fields with the original ones, their ratio (expressed in decibel (dB)) is displayed in Fig. IV.27 and Fig. IV.28. 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 scales 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 (as it was observed in Chapter III.3.1.1.1, see Fig. III.31). 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.

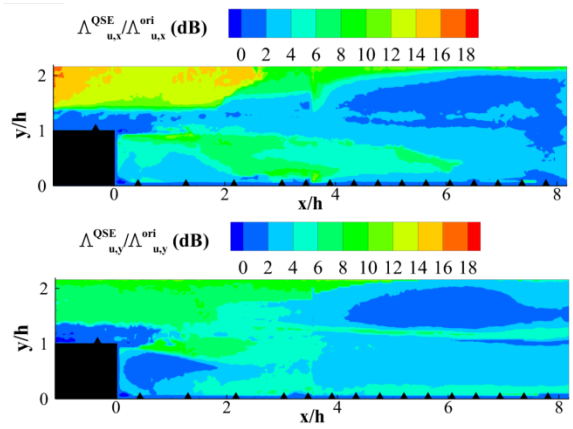


Fig. IV.29: QSE prediction to original turbulent spatial integral length scale for *u* ratio (top: calculated in streamwise direction; bottom: calculated in vertical direction)

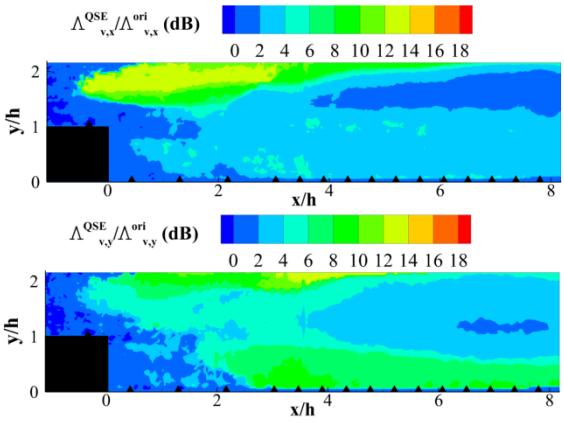


Fig. IV.30: QSE prediction to original turbulent spatial integral length scale for *v* ratio (top: calculated in streamwise direction; bottom: calculated in vertical direction)

Maps of the ratios of scales predicted by QSE on the original ones are plotted in Fig. IV.29 and Fig. IV.30. In overall ratios are decreased in the entire flow field. They are particularly lower in the boundary layer upstream of the step. That is interesting since it was shown in Chapter III that the QSE only improved the reconstruction accuracy and not the prediction accuracy in the BFS test case. In fact, maps of turbulent spatial integral length scales calculated from the training set or the validation set for the QSE prediction are almost identical. As for RMS quantities, the overfitting is not observed. A correct prediction of the turbulent spatial integral length scales is therefore not enough to conclude on the instantaneous prediction accuracy. At last, even if the estimated length scales are closer to the original ones using QSE, they are still increased. The QSE does not estimate accurately the turbulent spatial integral length scales either.

3.2. Linear Stochastic Estimation using velocity field sensors

Since the use of wall pressure sensors leads to a poor conservation of the turbulent spatial integral length scales, it was decided to investigate the impact of using velocity sensors in the flow on the conservation of these scales. For this study, eight grids of streamwise velocity sensors (extracted from the PIV measurements) are employed. The characteristics of each grid are summarized in Table IV.1. 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. IV.31 and Fig. IV.32 for sensors locations). Table IV.2 compares the determination coefficient of each configuration. It can be seen that determination coefficients up to 62% can be reached with the densest grid. With the coarsest grids, which contains 21 sensors (only 4 more sensors than there were wall pressure sensors in the previous case), a determination coefficient of 21% is reached, which is more than twice the one obtained with wall pressure measurements. The use of streamwise velocity information in the flow can clearly improve the estimation accuracy obtained by LSE.

For brevity, only figures of the longitudinal integral length scale of *u* are displayed for each grid. Fig. IV.31 and Fig. IV.32 show the maps of this integral length scales 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 prediction. For the grids of 462 and 240 sensors, the contours present a shape similar to the one in the top image of Fig. IV.23. When the sensor grids are coarsened, the contours become closer to the ones from LSE using pressure sensors. Also, the shortest integral length scale order of magnitude increases with the sensors spacing. Even with the finest grid, the integral length scales are not perfectly predicted and are increased in the entire domain. Therefore, they are also increased in the area of largest integral length scales

in the original data. These observations are also valid for the three other integral length scales $\Lambda_{u,y}$, $\Lambda_{v,x}$, $\Lambda_{v,y}$ (displayed in annex H.5, Fig. H.7 to Fig. H.12). In overall, the turbulent spatial integral length scales are overestimated in the entire field and in particular between the sensors where the correlations between the sensors and the velocity field are the lowest. This indicates a filtering of the length scales, the largest being better estimated than the shortest.

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

The approximate orders of magnitude of the shortest integral length scales estimated (found downstream of the step) are summarized in Table IV.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 shortest 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 these scales. But the smallest integral length scales seem to require sensor spacing of the order of magnitude of the turbulent scales.

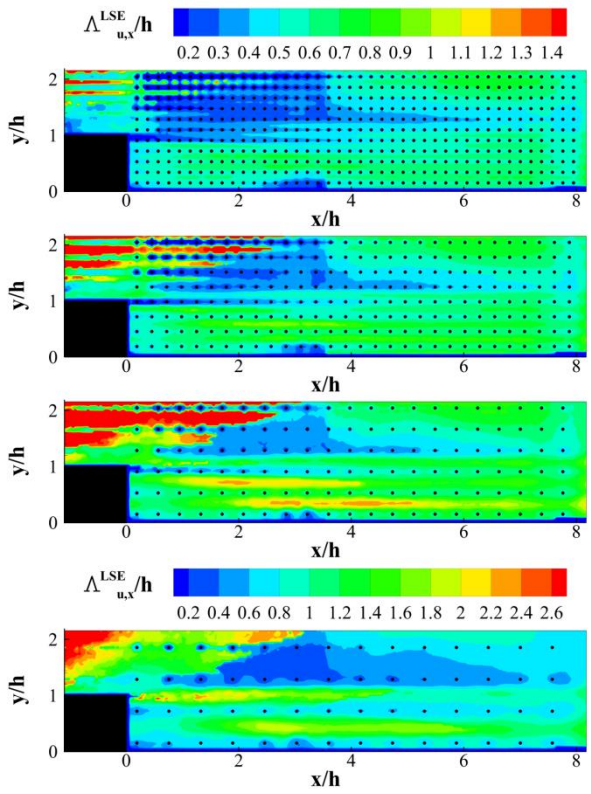


Fig. IV.31: Turbulent spatial integral length scales for *u* calculated in the streamwise direction for LSE using several grids of 462, 240, 126 and 56 sensors (from top to bottom)

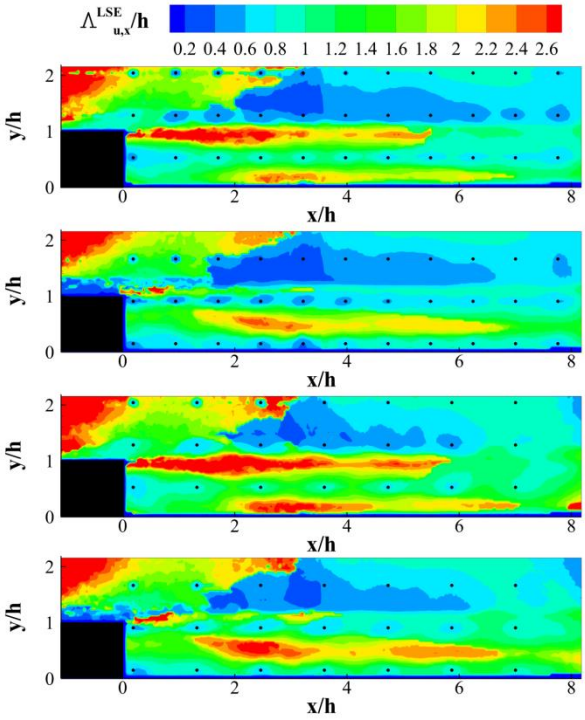


Fig. IV.32: Turbulent spatial integral length scales for *u* calculated in the streamwise direction for LSE using several grids of 33 and 21 sensors (from top to bottom)

3.3. Study through Proper-Orthogonal-Decomposition analysis

To complete the previous study, the estimation of the POD modes of the flow field through LSE is investigated. Details on the POD basis can be found in Chapter III.3.1.1.2. In addition, the POD convergence and its impact on the presented results are discussed in annex C. The determination coefficient $R^2_{\text{POD Single}}$ of the first thousand POD modes is plotted in Fig. IV.33 for the grids of 462, 240, 126 and 56 sensors. $R^2_{\text{POD Single}}$ for the first hundred POD modes is also plotted in Fig. IV.34 for the grids containing 33 and 21 sensors. In the present case, the general trend is that the higher the POD mode rank is, the lower the determination coefficient is. It shows that the higher POD modes are clearly less correlated to the sensors than the low POD modes. 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%.

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/λ in the spatial Fourier domain. Mathematically, the characteristic length λ of each POD mode *i* is defined by:

$$\arg\max_{\lambda\in\mathbf{R}^+}\left(\frac{\left|\iint H(\zeta,\xi)\widehat{\mathbf{\Phi}}_i(\zeta,\xi)\,d\zeta d\xi\right|^2}{\left|\iint \widehat{\mathbf{\Phi}}_i(\zeta,\xi)d\zeta d\xi\right|^2}\geq 0.95\right) \quad \text{(IV.1)}$$

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

$$H(\zeta,\xi)=\begin{cases}1 \text{ if } D(\zeta,\xi)<1/\lambda\\ 0 \text{ if } D(\zeta,\xi)\geq 1/\lambda\end{cases} \quad \text{(IV.2)}$$

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

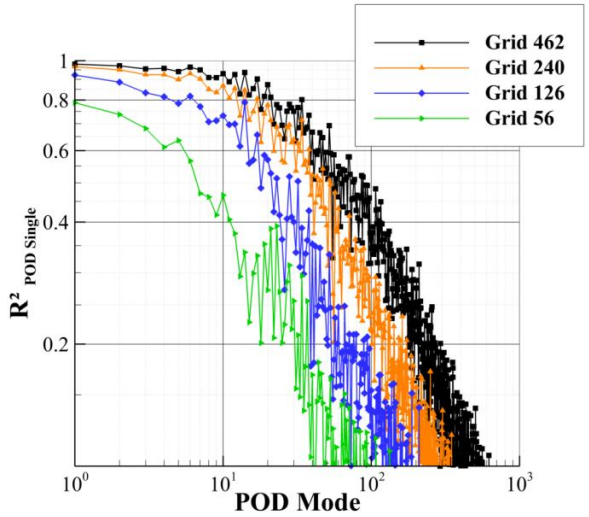


Fig. IV.33: Determination coefficient calculated for each POD mode separately on the validation set using LSE and several grids of streamwise velocity measurements

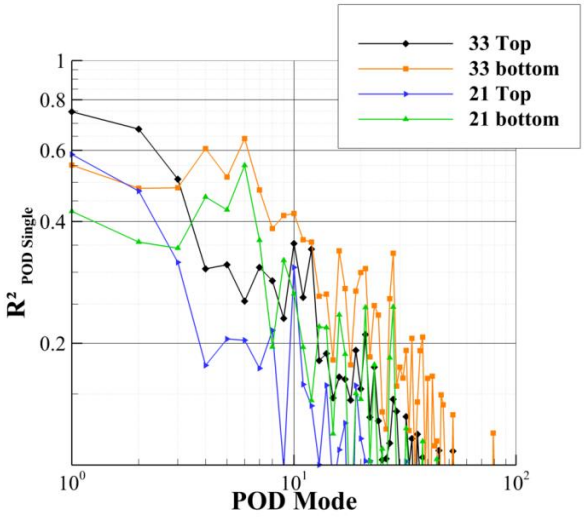


Fig. IV.34: Determination coefficient calculated for each POD mode separately on the validation set using LSE and several grids of streamwise velocity measurements

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. IV.35 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. Concerning POD modes which are not perfectly converged, λ obtained for the POD basis from the training set and from the first 3397 PIV snapshots of the training set were compared. The order of magnitude of λ as function of the POD mode rank is very close for both POD bases. In particular, it is true for POD modes of rank 30 to 70, which can be considered as converged for the first POD basis but not for the second one.

We now arbitrarily choose a threshold of 80%, for example, 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 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 turbulent 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.

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 arbitrarily choose a 10% threshold defining a cut-off rank for the POD modes which are summarized in Table IV.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 normalized by the sensor spacing d, that the characteristic length decreases when the grid is coarsening. This decrement is probably not the result of the lack of convergence of the higher order POD modes since it is already observed between POD modes 21 and 50 which are well converged (see annex C).

Furthermore, considering the convergence of the POD modes, it can be questionable to insert poorly converged POD modes into the model formed by the LSE. One may argue that the prediction of the entire flow field would be more accurate if these POD modes were disregarded. For brevity, let us consider only LSE using 462 streamwise velocity sensors. The determination coefficients of the validation set velocity field estimation is 61.9% (which corresponds to the use of all 6 793 POD modes). If only the first 650 POD modes (90% TKE) are used, then R^2 on the validation is 61.8%. Thus, the inclusion of the last 6 143 POD modes is unnecessary, but is also not harmful to the LSE model. Now, if only the first 70 POD modes, which can be considered as converged, are estimated, then R^2 falls to 56.3%. Therefore, estimating not converged POD modes, still improves the prediction of the entire velocity field outside the training set.

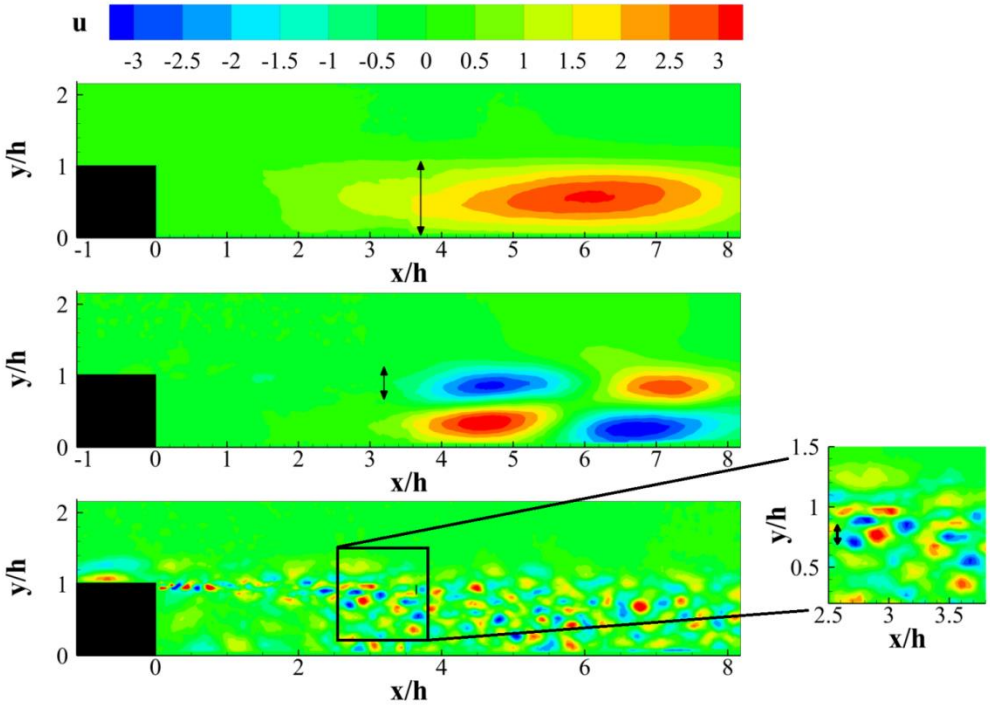


Fig. IV.35: Streamwise component of the velocity POD modes 1, 7 and 400 (from top to bottom). Black arrows illustrate the characteristic length of the mode.

In Fig. IV.33 and Fig. IV.34, it can be seen that the determination coefficient rapidly decreases with the POD mode rank. For the grids 462 to 126, the POD mode prediction accuracy

quickly deteriorates for POD modes of ranks higher than 30. For the other grids, the deterioration starts 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. III.37). 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 0.6 h.

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

Table IV.3: Cut-off rank of predicted POD modes and characteristic length associated for LSE using several grids of streamwise velocity sensors.

Looking more closely to Fig. IV.34, 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.

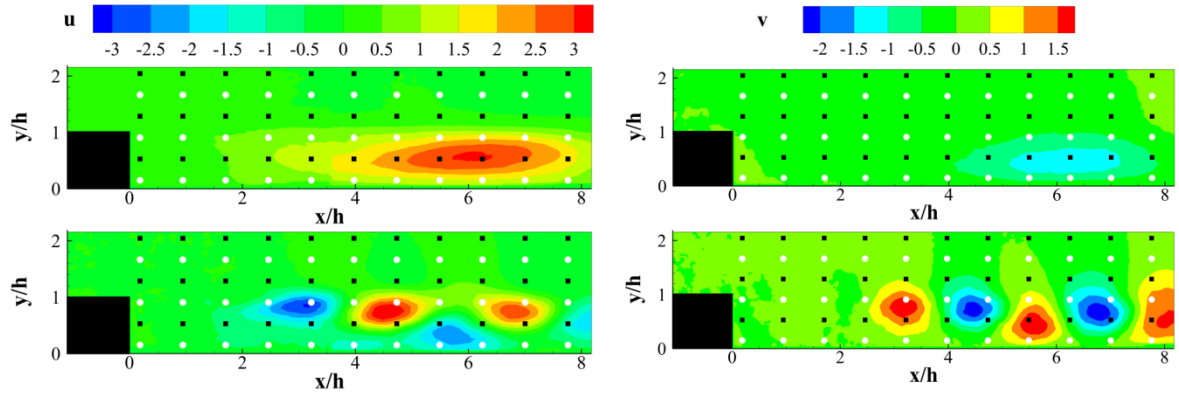
3.4. Sensor locations impact

Differences in the estimation accuracy of the POD modes between two grids, with the same number of sensors and the same spacing, induce 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 turbulent spatial length scales 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 in the flow through SE.

In 2006, Cohen et al. [63] proposed a heuristic method to place the sensors in order to estimate POD coefficients using LSE. Their method uses a POD of the possible sensors and they advocated locating the sensors at the extrema of the sensor POD modes. In our case, when using velocity measurements from the PIV data as sensors, they should therefore be located at the extrema of the velocity POD modes.

The first and sixth modes are displayed in Fig. IV.36 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. IV.34 than 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 component, than the grid 33 "top". Extrema of the POD modes seems to be indeed a better location for the sensors. To correctly estimate a POD mode, it appears to be important to dispose of sensors near the extrema of the mode.



a) Streamwise velocity component

b) Vertical velocity component

Fig. IV.36: Streamwise (a) and vertical (b) component of the velocity POD modes 1 and 6. Black squares corresponds to the 33 "top" sensor grid and the white circles to the 33 "bottom" grid.

As Fig. IV.33 and Fig. IV.34 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 in the present situation. Table IV.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 that the determination coefficient decreases with the POD mode rank and that more sensors are necessary to estimate higher POD modes with the same level of R^2. This confirms the difficulty which arises when trying to estimate small scale structures from a limited number of measurements through LSE. 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. The positioning of sensors will be addressed more closely in Chapter V.

POD mode	R^2 using only one sensor
1	37%
2	30%
5	20%
10	17%

Table IV.4: Determination coefficient of the prediction of several POD modes when using only one streamwise velocity sensors located at the extremum of the POD mode

3.5. Impact of Multi-Time-Delay method

3.5.1. Multi-Time-Delay Linear Stochastic Estimation using wall pressure measurements

The best situation tested in Chapter III.3.1.2 is here considered. As previously shown, improvements in accuracy were found to be marginal (less than 1% for the determination coefficient calculated on the validation set). Ratio of predicted (by MTD-LSE) to original spatial integral length

scales are shown in Fig. IV.37 for the streamwise velocity and in Fig. IV.38 for the vertical velocity. These figures can be compared with Fig. IV.27 and Fig. IV.28. One can observe that, in overall, the ratio is only slightly decreased when MTD-LSE is used instead of LSE.

The determination coefficient of the first 90 POD modes predicted by MTD-LSE-POD (using 75 delays every 1×10^{-2} $T_{shedding}$) is plotted in Fig. III.42. It was shown that the MTD-LSE-POD prediction is slightly better than the LSE-POD. But both perform badly for any POD modes, with R^2 lower than 25%. The cut-off rank is about 10 in both situations, but the MTD-LSE-POD is able to recover some POD modes, of ranks between 20 and 30, with a R^2 higher than 10%. To give an order of idea, the 10th POD mode has a characteristic length of about 0.6 h. Therefore, even when using MTD-LSE, structures shorter than half the step are not correctly predicted.

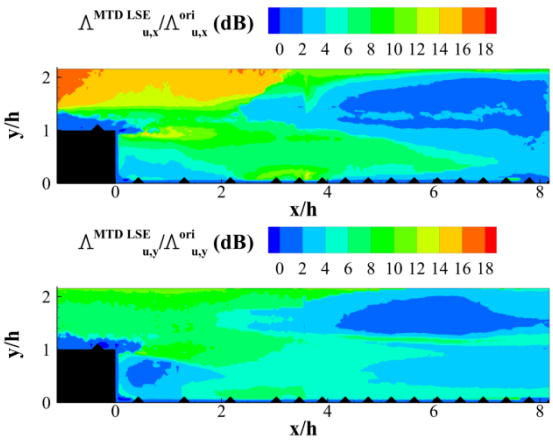


Fig. IV.37: MTD-LSE prediction to original turbulent spatial integral length scale for *u* ratio (top: calculated in streamwise direction; bottom: calculated in vertical direction)

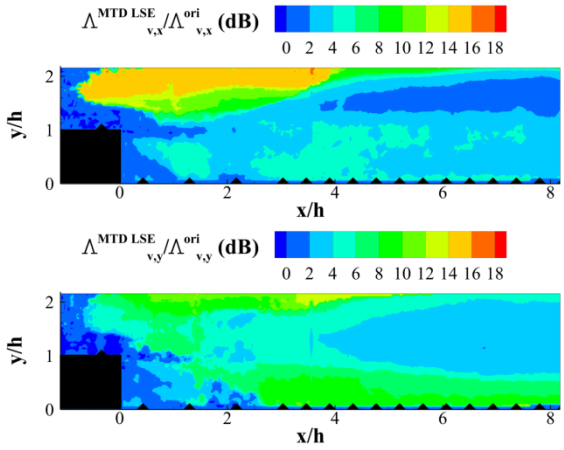


Fig. IV.38: MTD-LSE prediction to original turbulent spatial integral length scale for *v* ratio (top: calculated in streamwise direction; bottom: calculated in vertical direction)

3.5.2. Multi-Time-Delay Linear Stochastic Estimation using streamwise velocity measurements

To perform a MTD-LSE using streamwise velocity measurements as conditional events, it is necessary to dispose of measurements sampled at frequency higher than the shedding frequency (about 100 Hz, St = 0.1). Thus, the use of data extracted from the low speed PIV is not possible and results from the high speed PIV are utilized. The LSE using 20 streamwise velocity measurements from the high speed PIV database do not lead to any overfitting (the difference of R^2 between the validation set and the training set is less than 1%). The choice of locations for the 20 measurements will be addressed precisely in Chapter V. Once more the best MTD-LSE configuration among those tested is considered. In this configuration, 10 delays taken every 5×10^{-2} $T_{shedding}$ for each sensor are used. The determination coefficient on the validation set is about 39% using LSE and 45% using MTD-LSE (with Tikhonov regularization). Thus, the improvement is more significant than when using the wall pressure sensors.

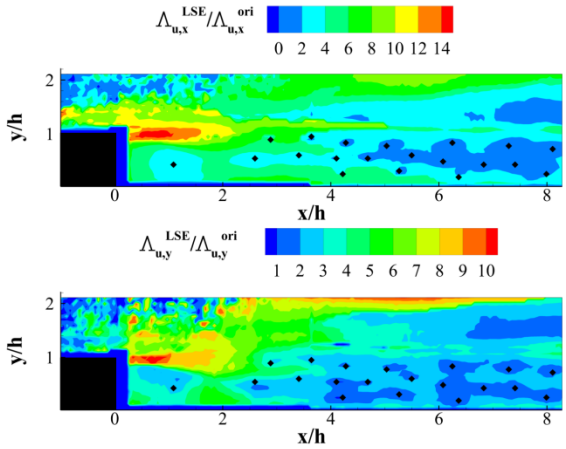


Fig. IV.39: LSE prediction to original turbulent spatial integral length scale for *u* ratio. Black dots show the position of the streamwise velocity sensors (top: calculated in streamwise direction; bottom: calculated in vertical direction)

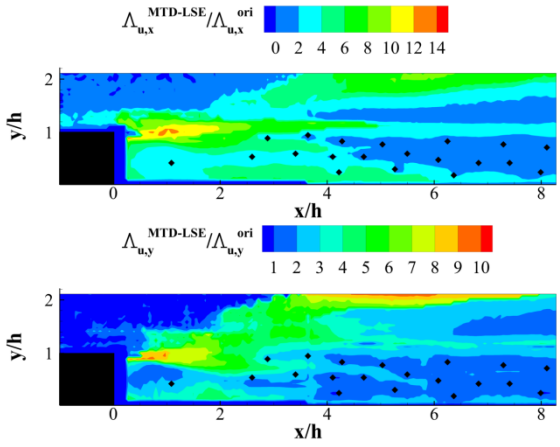


Fig. IV.40: MTD-LSE prediction to original turbulent spatial integral length scale for *u* ratio. Black dots show the position of the streamwise velocity sensors (top: calculated in streamwise direction; bottom: calculated in vertical direction)

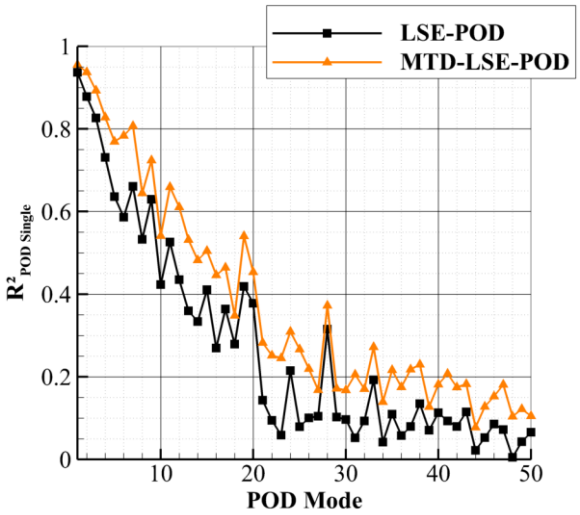


Fig. IV.41: Determination coefficient calculated for each POD mode separately on the validation set using LSE and MTD-LSE with 20 streamwise velocity measurements

Fig. IV.39 and Fig. IV.40 compare ratio of predicted to original integral length scales (for the streamwise velocity) obtained in LSE and MTD-LSE case. The ratio is lower in the MTD-LSE case and in overall MTD-LSE is able to recover smaller length scales than the LSE. $R^2_{POD\ Single}$ of the first 50 POD modes predicted by LSE-POD and MTD-LSE-POD are shown in Fig. IV.41. The prediction of each POD mode is clearly improved by MTD-LSE-POD. $R^2_{POD\ Single}$ increases up to 20% for some POD modes. Using a cut-off at 10% as previously, LSE-POD is unable to predict POD modes of ranks higher than 30 ($\lambda \approx 0.44\ h$), while MTD-LSE-POD is able to predict POD modes up to rank 50 ($\lambda \approx 0.32\ h$). Therefore, LSE is able to estimate structures larger than about one half the step height and MTD-LSE larger than one third of the step height without requiring more sensors.

3.6. Summary

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 sensor 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. At last, it was demonstrated that the use of Multi-Time-Delay LSE (MTD-LSE) could improve the estimation of the spatial integral length scales.

4. Spatial and temporal pre-filtering impact on the Linear Stochastic Estimation accuracy

It was shown in previous parts that SE seems to filter the original velocity field both temporally and spatially. The goal of this section is then to keep studying this filtering effect but also to investigate the possibility to improve the estimation accuracy by filtering the data used by the SE. Following the study of the length scale estimation, spatial filtering is first considered. In this section, only the backward facing step flow is used.

4.1. Spatial filtering

Previously, LSE was shown to only be able to correctly estimate the first POD modes of the backward facing step velocity field. These POD modes were also the ones corresponding to the longest length scales. Therefore, it is legitimate to compare the LSE estimations with spatially filtered original fields. In addition, the opportunity to spatially filter the training fields is investigated.

The spatial filter used here is a 2D Gaussian spatial filter which is a low pass filter (thus only the longest lengths are conserved). The determination coefficient of the LSE estimation using the 17 pressure sensors and unfiltered training fields is displayed in Fig. IV.42. The reference for the calculation is the original fields after they have been spatially filtered. Thus, the determination coefficient is plotted as function of the cut-off length L_c used to filter the original velocity fields. R^2 goes by a maximum when the cut-off length is increased. The estimated fields are found to be closer to the original fields when they have been spatially filtered with a cut-off length about 0.84 h. From the results of Chapter IV.3, it is logical that spatially filtered fields, where small length scale structures have been filtered out, are getting closer to the estimated fields. It was also observed that the first two POD modes were the best reconstructed and predicted. The structures contained in these two POD modes are of the size of the step which is the order of magnitude of the “optimal” cut-off length. Original fields filtered at higher cut-off lengths then diverge from the estimated fields and R^2 decreases. It is due to the fact that the filtered fields used for reference do not contain structures that have been estimated because they have been filtered out.

Determination coefficients calculated from spatially filtered original fields for LSE estimation using 20 streamwise velocity measurements (the same that were used in Chapter IV.3.5.2) are plotted in Fig. IV.43. Similarly to the case where 17 pressure sensors are used, R^2 goes by a maximum when the cut-off length L_c increases. However, the "optimal" cut-off length is smaller. It is about 0.6 *h* instead of around 0.84 *h*. Therefore, from the previously given explanation, it induces that estimated fields, for the LSE using streamwise velocity sensors, contain structures of shorter length than when pressure sensors are used. That is consistent with what was shown in Chapter IV.3.5.2. It was found that using this set of streamwise velocity measurements, POD modes up to 30 were predicted with R^2 higher than 10. The characteristic length of the 30th POD mode is about 0.44 *h* which is clearly shorter than the characteristic length of the 2nd POD mode (~ 1 *h*).

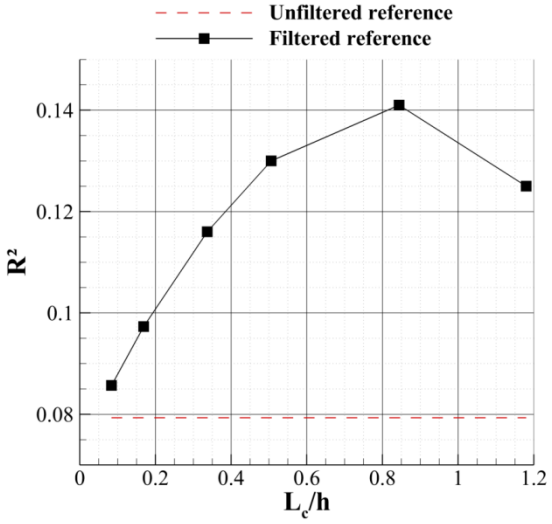


Fig. IV.42: Determination coefficient (validation set) of unfiltered LSE (with 17 pressure sensors) using spatially filtered original fields as reference. (The use of unfiltered original fields as reference is also plotted, red dashed line)

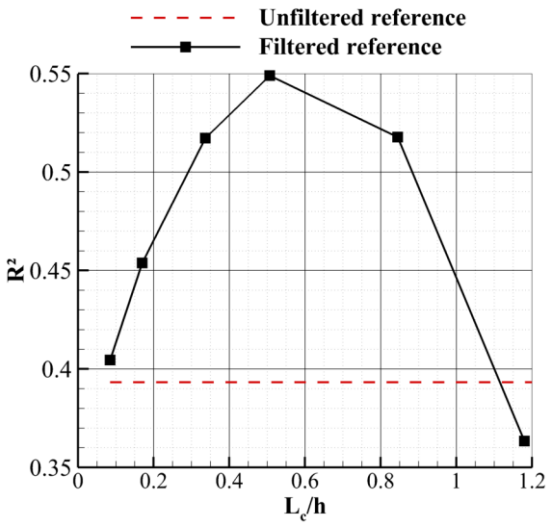
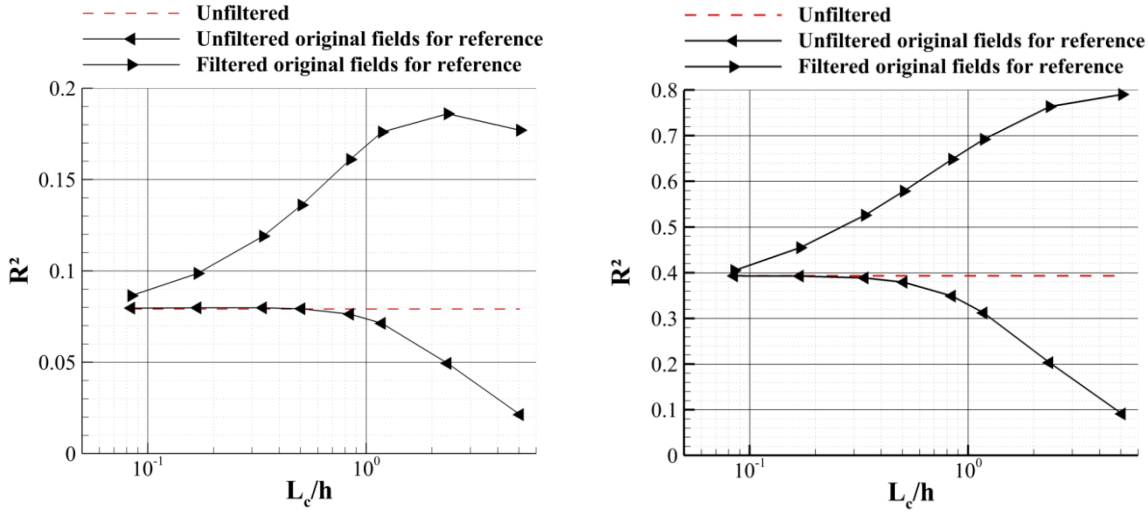


Fig. IV.43: Determination coefficient (validation set) of unfiltered LSE (with 20 streamwise velocity sensors) using spatially filtered original fields as reference. (The use of unfiltered original fields as reference is also plotted, red dashed line)

The filtering of the training set velocity fields is now investigated. Fig. IV.44 shows the evolution of R^2 with the cut-off length when the unfiltered and filtered validation set fields are used as reference (LSE using 17 pressure sensors and 20 streamwise velocity sensors are shown). It is clear that spatially pre-filtering the training set fields does not improve the estimation of the unfiltered original fields. The curves clearly highlight a cut-off length of the structure estimated. Indeed, as the filter cut-off length increases, first the determination coefficient remains unchanged, and then at some point, it quickly decreases. That means that structures of lengths lower that the cut-off length (determined by the point where the decreasing starts) were not in the estimated fields. Therefore, filtering them from the training fields does not impact the estimation overall accuracy. But, if too large structures are filtered out of the training fields, and that these structures were estimated, then they are filtered out of the estimated fields. The estimated fields then become lesser representative of the original fields and R^2 decreases. From these figures, LSE using 17 pressure sensors seems unable to estimate structure of length lower than the step height, and LSE using 20 streamwise velocity sensors lower than 0.5 *h*. These values are perfectly consistent with the cut-off characteristic length of the POD modes (when a 10% threshold was considered). The study of such curves is obviously another way to determine the length of the smallest structures estimated without using the POD decomposition and the characteristic lengths associated.

When R^2 is calculated using the filtered fields as reference, it first increases with the cut-off length, then reaches a maximum and decreases (the decreasing part of the curve is not seen for the 20 streamwise velocities case, but the behavior displayed by the curve indicates that R^2 will at least reach a threshold). It therefore seems possible by spatially filtering the training fields to select structures (according to their length) and to slightly improve their estimation. Indeed, R^2 reaches higher values when the training fields are filtered than when they are not (and that R^2 is calculated from filtered original fields). The values obtained with the filtered fields as reference are higher in Fig. IV.44 (training fields are filtered) than in Fig. IV.42 and Fig. IV.43 (training fields are not filtered).



a) LSE using 17 pressure sensors b) LSE using 20 streamwise velocity sensors

Fig. IV.44: Determination coefficient as function of the cut-off frequency of the spatial low-pass filter applied to the training fields

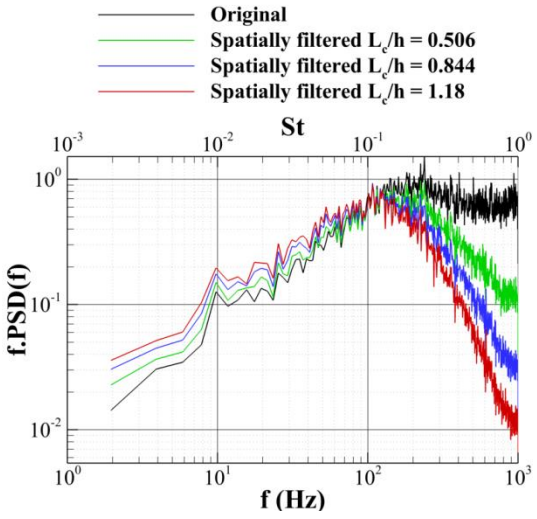


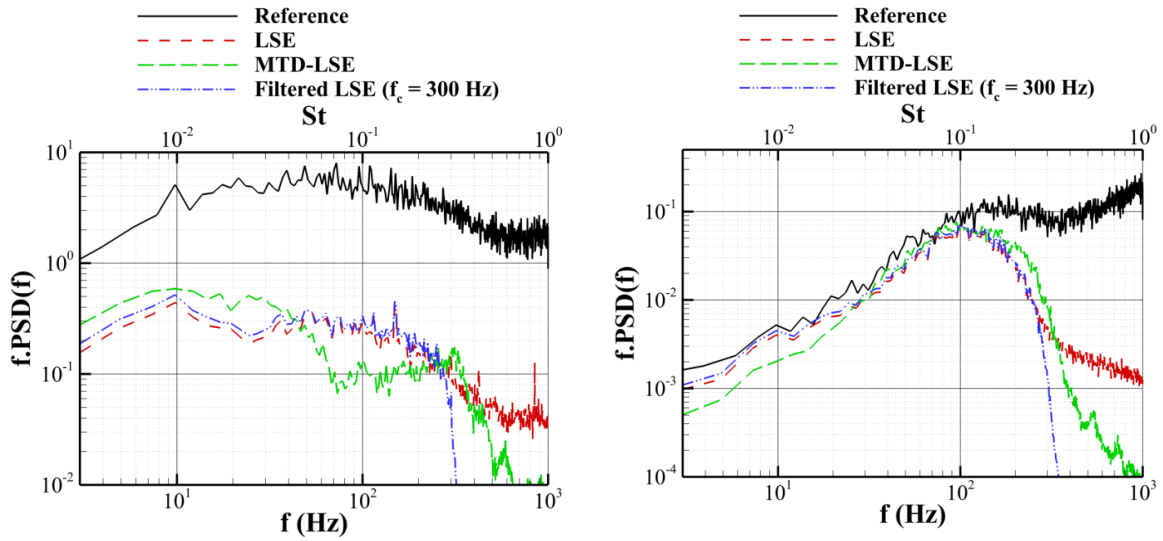
Fig. IV.45: Power Spectral Density of the vertical velocity at the same location as in Fig. IV.47 for several spatial filters of the velocity field

Vertical velocity PSD, at the same locations as in Fig. IV.47, obtained after spatially filtering the original fields, are displayed in Fig. IV.45. As it is expected, spatially filtering the original fields also impact the frequency content. The low frequencies are conserved (for this location the low frequencies are amplified but it is not the case in most of the velocity field). On the contrary, the

high frequencies are cut. This is expected since high frequencies are linked to small scale structures. The cut-off frequency depends on the cut-off length used. The larger the cut-off length is, the lowest the cut-off frequency is. Also the high frequencies are more and more attenuated by increasing the cut-off length. Thus, the spatial filtering allows the selection of frequencies around the shedding which can explain the improvement of R^2 when using the spatially filtered fields as reference. Also if L_c is too high, low frequency content, that are estimated, are filtered out of the reference field, which decreases R^2. The temporal filtering of the data is now investigated.

4.2. Temporal filtering

4.2.1. Spectral study



a) Point 1 b) Point 3

Fig. IV.46: Power spectral density of several estimations of the validation set compared to the original spectrum (streamwise velocity)

It was shown in Chapter III.3 that reconstruction and prediction of the velocity field downstream the backward facing step using SE with 17 wall pressure sensors were poorly accurate. Determination coefficient R^2 (from the validation set) ranges from 6.9% (QSE) to 9.3% (MTD-LSE). As done for the OAT15A test case, the spectral content of SE prediction is quickly investigated. Only the LSE and MTD-LSE results (from the low speed PIV) are discussed here, and the high speed PIV measurements are used to compute the reference spectra. To be used as reference, the high speed PIV spectra are corrected to match the RMS of the low speed PIV at the considered point. Spectra of the prediction are obtained by estimating the velocity field at 10 kHz (using the pressure measurements). In the MTD-LSE case, 75 delays every 10^{-2} $T_{shedding}$ are used, which corresponds to the best situation obtained while testing several MTD-LSE configurations. Power Spectral Density (PSD) of the streamwise velocity at two locations (one in the recirculation bubble and one above the recirculation) are displayed in Fig. IV.46.a) and Fig. IV.46.b) respectively. The extracted location in the recirculation bubble is denoted as point 1 and the one above as point 3. PSD of the vertical velocity at one location (in the shear layer) are shown in Fig. IV.47 (the location is identified as point 2 in Fig. IV.49). From these PSD LSE and MTD-LSE results, one can conclude that both methods are unable to match the original spectrum. But, as observed in the OAT15A test case, low frequencies are better recovered. That is clear for point 2 and point 3. For these points, the spectral content predicted clearly falls from around 200 Hz to 400 Hz. For point 1 disparities between estimated and original spectra seems also more important for the high frequencies than for

the low ones. The MTD-LSE performs a little bit better for frequencies around 200 Hz to 400 Hz. Due to the use of the Tikhonov regularization the high frequency content is, however, even more attenuated than in the LSE case.

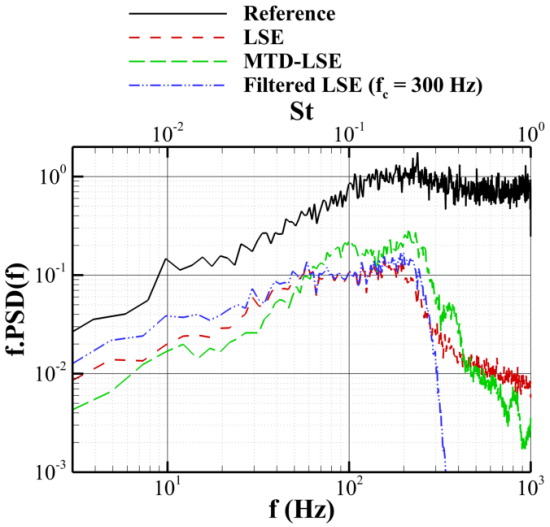


Fig. IV.47: Power spectral density of several estimations of the validation set compared to the original spectrum (vertical velocity at point 2)

From these observations, the opportunity to pre-filter temporally the pressure sensors and/or the field to estimate was investigated. First, only the filtering of the pressure sensors was studied because it is not possible to filter, at frequencies higher than 5 Hz, the velocity fields obtained from low frequency PIV. In addition, it was shown in Chapter IV.1 that the noise contained in the conditional events directly deteriorates the estimation quality. As one can fairly expect the noise to be more prominent at high frequencies, low-pass filtering of the conditional events can delete part of the noise and could improve the estimation.

4.2.2. Filtering of conditional events alone (low speed PIV database)

The temporal filter used is a low-pass 8th-order Butterworth filter. The filter is applied forward and backward in time to keep the phase unchanged. Validation set R^2 as function of the cut-off frequency is displayed in Fig. IV.48. The R^2 corresponding to the situation where the pressure sensors are unfiltered is also plotted for comparison (dashed line). From these data, one can see that the estimation can be slightly improved by pre-filtering the conditional events alone. The cut-off frequency, obviously, impacts the estimation quality. As the cut-off frequency (f_c) increases, the determination coefficient goes by a maximum between f_c = 300 and f_c = 400 Hz. Cutting lower frequencies leads to a deterioration of the original unfiltered velocity fields estimation. It can be explained by the fact that a too important part of the velocity field is no more estimated and is filtered out. On the contrary, when cutting at higher frequencies, the additional frequencies estimated act like a noise that deteriorates the estimation and R^2 decreases. The value of the optimal cut-off frequency is also coherent with what was just observed, where the LSE (and MTD-LSE) were shown to cut frequencies higher than 400 Hz. Also the PSD of the pressure signals showed that most of the energy they contained is at frequencies smaller than 400 Hz (see Fig. II.14). It is also important to notice the strong deterioration of the reconstruction when cutting at lower frequencies than 300 Hz, compared to the weak deterioration when higher frequencies are kept. This fact suggests that most of the estimated content is a low frequency one which is consistent with observations made from PSD.

Three predicted fields are displayed in Fig. IV.49. They correspond to the same instant of time. One is obtained using LSE and the unfiltered pressure sensors, the other two using LSE and pressure sensors low-pass filtered at f_c = 400 Hz and f_c = 200 Hz (the original field is displayed in Fig. III.29). The overall shape of the predicted field is not modified by the filtering at 400 Hz. However, the fluctuation levels are slightly changed. More particularly the range of fluctuations estimated increases (and it is observed generally over different instants of time). If the pressure sensors are filtered with a smaller f_c, the structures are modified and the range of fluctuations decreases, which explains the decrease of R^2. A part of the velocity field correctly predicted has been filtered out.

The PSD of the prediction when the pressure sensors have been low-pass filtered at f_c = 300 Hz are plotted in Fig. IV.46 and Fig. IV.47. These figures show frequencies lower than 100 Hz are better recovered using the filtered conditional events than the unfiltered one (when using LSE). The cut-off frequency is clearly seen with a sudden fall at 300 Hz. Thus frequencies higher than f_c are indeed not estimated.

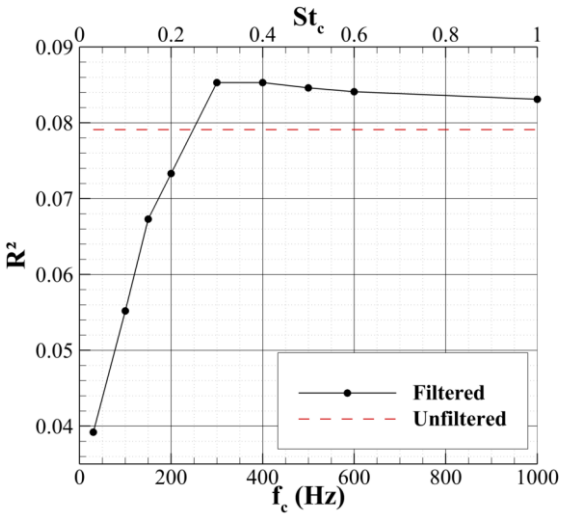


Fig. IV.48: Determination coefficient as function of the cut-off frequency of the low-pass filter applied to the pressure sensors (LSE from low speed PIV data and pressure sensors)

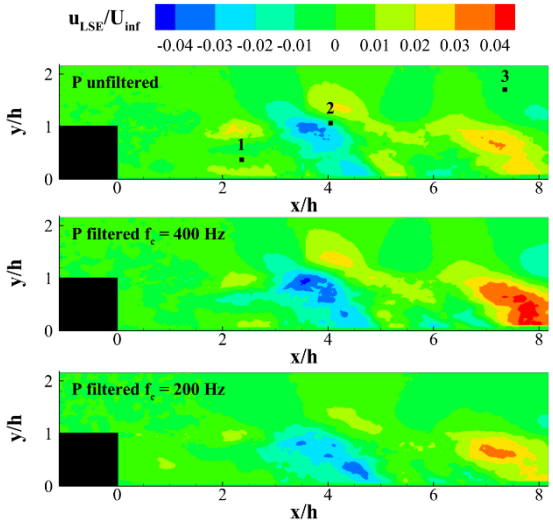


Fig. IV.49: Predicted streamwise velocity field comparison between several LSE using filtered and unfiltered pressure measurements

At last the possibility to band-pass filter the pressure data is also quickly investigated. From the pressure signal PSD (see Fig. II.14.a)), it is seen that most of the energetic content is contained between 30 Hz and 400 Hz. Therefore, the pressure signal is band-pass filtered between these two frequencies. The determination coefficient of the prediction is then 5.73%. The prediction quality is lower than in the unfiltered case but also than in most of low-pass filtered cases. Therefore, even if the low frequency content of the pressure sensors seems to have a negligible energy compared to the energy contained between 30 Hz and 400 Hz, the estimation of frequencies lower than 30 Hz represents an important part of the overall estimation.

4.2.3. Filtering of the conditional events and training fields (high speed PIV database)

If, in our case, the SE reconstruction and prediction accuracy can indeed be improved through the temporal pre-filtering of the conditional events, the improvement is not significant. Thus the impact of filtering the fields to estimate is now investigated. For this study only high speed PIV

data are used, since it is necessary to dispose of velocity fields samples at high frequency. It was previously shown that no overfitting was observed using LSE on the high speed PIV database. So the evolution tendencies are expected to be still representative. Also no pressure measurements are available in this situation. Instead streamwise velocities, extracted in 20 locations in the flow field, are used (as in Chapter IV.3.5.2).

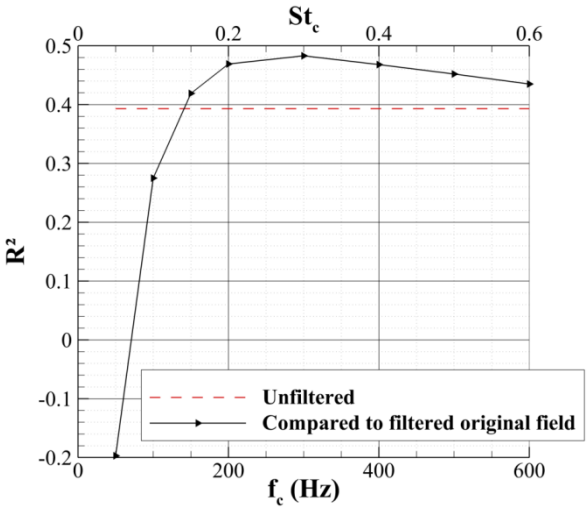


Fig. IV.50: Determination coefficients as function of the cut-off frequency of the low-pass filter applied to the original fields used as reference (LSE from unfiltered data)

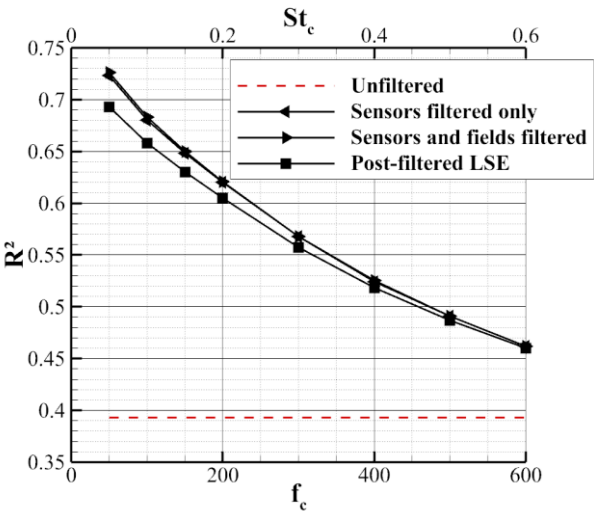


Fig. IV.51: Determination coefficient as function of the cut-off frequency of the low-pass filter applied to the sensors and training fields (LSE from high speed PIV data and streamwise velocity measurements; original filtered field used as reference; Post-filtered LSE corresponds to LSE from unfiltered measures and training field which results are then filtered)

The LSE prediction (using unfiltered sensors and training field) is first compared with low-pass filtered original velocity fields for the calculation of R^2. The determination coefficient calculated using these filtered fields as reference is plotted in Fig. IV.50. Values of R^2 reaches a maximum at f_c = 300 Hz. Since only the reference, used to calculate R^2, changed, it indicates that the estimated content is mainly a low frequency one, more particularly, under 300 Hz. This observation is consistent with previous ones. Another interesting way to identify the quality of the estimation, according to the spectral content, is to look at the determination coefficient obtained from the filtered part of the estimated fields compared to the filtered original fields. This R^2 is plotted in Fig. IV.51 (curve with square symbols). It shows a decrease of the prediction quality as more and more high frequencies are taken into account. Thus, it confirms that the low frequencies are the best predicted.

4.2.3.1. Filtering of the conditional events alone

The same analysis as previously is now conducted and only the sensors data are low-pass filtered. The validation set determination coefficient as function of the cut-off frequency is plotted in Fig. IV.53. Once more the R^2 obtained without filtering the sensors is also plotted. The trends observed previously are perfectly reproduced here. While decreasing the cut-off frequency, the determination coefficient first increases, goes by a maximum around 400 Hz and then decreases rapidly. It is interesting to note that the cut-off frequency, leading to the best R^2, is of the same order of magnitude than in the previous case whereas the conditional events have been changed in nature

and number (streamwise velocity are used instead of wall pressure). This observation suggests that this "optimal" cut-off frequency is not dependent of the conditional events used for the LSE, but directly linked to the flow that is estimated. The PSD of three of the twenty streamwise velocity data (locations shown in Fig. IV.54), used as conditional events, are displayed in Fig. IV.52. The Kolmogorov's law is well observed and whatever the velocity data considered the cascade starts about 200 Hz. In addition, it is possible to see a plateau or rebound around 500 Hz that is symptomatic of noise contained in the PIV measures. This observation may explain that the "optimal" cut-off frequency is found around 400 Hz.

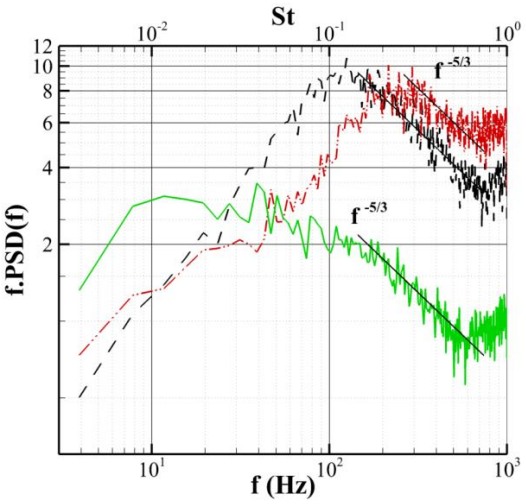


Fig. IV.52: Power spectral density of three streamwise velocity data used as conditional events for the LSE (Point 1: red plot, Point 2: black plot, Point 3: green plot)

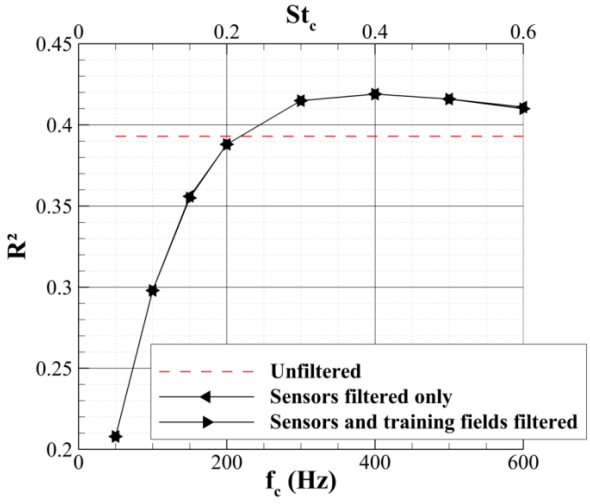


Fig. IV.53: Determination coefficient as function of the cut-off frequency of the low-pass filter applied to the sensors and training fields (LSE from high speed PIV data and streamwise velocity measurements; original unfiltered field used as reference)

4.2.3.1. Filtering of both conditional events and training fields

The impact of filtering the fields used to train the LSE is now studied. First, the determination coefficient obtained for LSE with the conditional events filtered and the training fields filtered is compared, in Fig. IV.53, with the case where only the conditional events have been filtered. The reference for the calculation of R^2 is of course the unfiltered velocity fields in this figure. The two curves match each other almost perfectly. If the conditional events are filtered, then using training fields filtered at the same cut-off frequency does not change the estimation by LSE. This should be expected and can be simply explained.

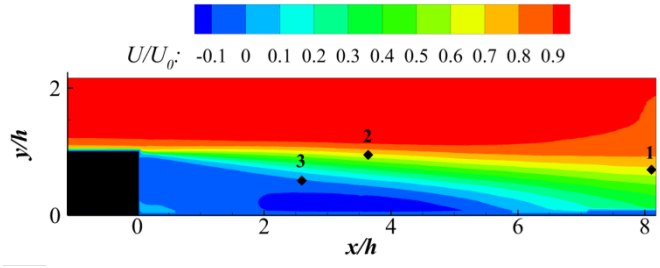


Fig. IV.54: Average streamwise velocity field with the locations of the 3 streamwise velocity measurements which PSD are displayed in Fig. IV.52

To simplify the discussion, let us consider the case of a perfect low-pass filter. Let us suppose that the training fields and conditional events are both low-pass filtered. We denote f_c the lowest cut-off frequency applied to either the training fields or the conditional events. Then the cross-spectral density of any conditional event with any point of the field (Φ_{E-U}) verifies:

$$\Phi_{E-U}(f) = 0 \text{ if } f > f_c \tag{IV.3}$$

From the Wiener-Kinchine theorem and using the inverse Fourier transform we have:

$$\langle E(t)U(t)\rangle = \int_0^\infty \Phi_{E-U}(f)\, df \tag{IV.4}$$

Therefore, if the conditional events or the training fields are low-pass filtered with a cut-off frequency f_c, the previous integral range from 0 to f_c and the covariance between them is equal to the covariance in the case where both data have been low-pass filtered at f_c. Thus when only the conditional events are filtered, the system of equation leading to the LSE coefficients is equivalent to the one obtained if the training fields are also filtered.

In Fig. IV.51, R^2 calculated with the low-pass filtered original fields for reference, in the case where the sensors have been filtered, is displayed. The case where both the sensors and the training fields have been filtered is also plotted and matches the case where only the sensors are filtered, which is consistent with the previous discussion. In these situations, R^2 decreases and tends toward the unfiltered value. Thus, R^2 follows the same trend as when the LSE, obtained from unfiltered conditional events and training fields, is post-filtered and then compared with filtered original fields. The R^2 values are slightly better from the LSE using filtered sensors but the differences reach only 4% for f_c = 50 Hz. This is consistent with previous observations showing that filtering the conditional events leads to a slight improvement of the prediction.

4.2.3.2. Filtering of the training fields alone

At last, the impact of filtering the training fields alone is studied. The determination coefficient evolutions, as function of the cut-off frequency, calculated using the unfiltered original fields for reference (left triangle curve) and using the filtered original fields for reference (right triangle curve), are displayed in Fig. IV.55. Filtering the training field does not lead to any improvement of the estimation of the unfiltered fields. The left triangle curve always remains under the value of R^2 where no data are filtered and tends toward this value when f_c increases. Looking at the estimation quality of the filtered part of the original fields (right triangle curve), a best situation is found for f_c = 150 Hz. The interpretation of this situation is more complicated than previous cases. Indeed, in this case the estimation used unfiltered conditional events, thus only the matrix of the covariance between the conditional events and the fields to estimate is modified. Therefore, even if at the training only low frequencies of the fields are taken into account, higher frequencies can be estimated.

This may explain that R^2 goes by a maximum and does not simply increase when the cut-off frequency decreases. Higher frequencies are estimated and deteriorate the estimation of the low-pass filtered original fields. This is confirmed by the PSD of the vertical predicted velocity signal at the same location as in Fig. IV.47, which are plotted in Fig. IV.56. Two cut-off frequencies are considered 50 Hz and 150 Hz. The PSD of the filtered original signal are also plotted. Both reconstructed signals go by a peak around 100 Hz, therefore when the original signal is cut at 50 Hz, proportionally a more important part of the predicted spectrum is contained in frequencies that have been filtered out. In addition, the R^2 values, achieved by filtering the training fields and using the filtered original fields as reference, remain lower than those obtained in the case where the conditional events are filtered.

In overall it is possible to slightly improve the reconstruction and prediction accuracy by filtering the conditional events used for the LSE. However, the improvement is limited. On the contrary, filtering the training fields does not increase the quality of the estimation. This study also highlighted again that the low frequency content is the best estimated.

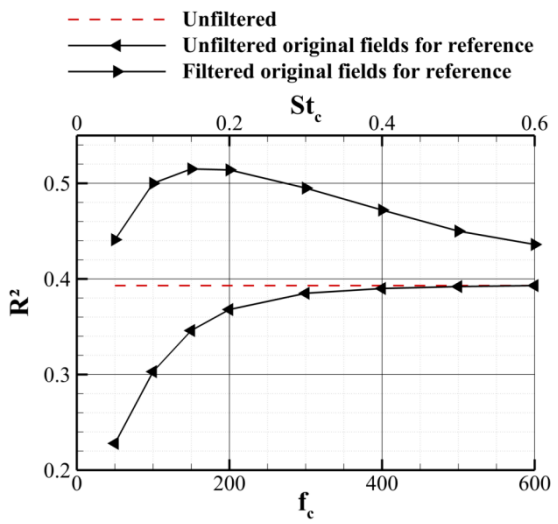


Fig. IV.55: Determination coefficient as function of the cut-off frequency of the low-pass filter applied to the training fields (LSE from high speed PIV data and streamwise velocity measurements)

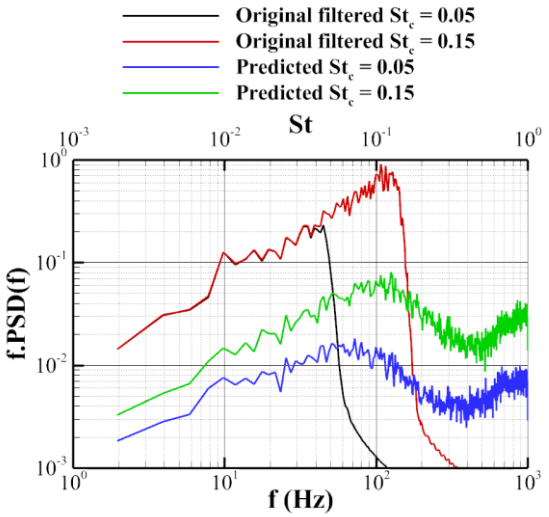


Fig. IV.56: Power Spectral Density of the predicted vertical velocity compared to filtered original signal (LSE using filtered training fields)

4.3. Summary

The impact of temporally pre-filtering the conditional events used for the LSE with a low-pass filter has been investigated. The results showed that such a pre-filtering could improve the reconstruction and prediction accuracy. Indeed, low-pass pre-filtering of the conditional events prevents the estimation of high frequencies that act like a noise on the estimate. The improvement is however very limited. When the conditional events are filtered, it has been proven that pre-filtering the training fields, at a higher cut-off frequency than the conditional events, does not bring any modification. Through this study, the fact that the low-frequencies are the best estimated was also highlighted. Thus selecting a priori these frequencies and targeting their prediction leads to a better prediction quality. The impact of temporally pre-filtering the training fields alone has also been explored. In this situation, it is not possible to improve the estimation of the unfiltered fields.

In a similar manner as with the temporal pre-filtering, spatial pre-filtering of the training fields did not lead to an improvement of the estimation of unfiltered fields. It is however possible

to select high scale structures that are then better reconstructed and predicted by the LSE. The study also highlights a clear cut-off of the LSE on the length of the estimated structures which confirms the observations of previous parts.

5. Impact of Reynolds number

The impact of varying the Reynolds number for a same flow configuration is here investigated. In this part, the flow around the wall-mounted cube is used and in order to conduct spectral studies, the high speed PIV database is mainly employed.

5.1. Determination coefficient evolution

Determination coefficients obtained using LSE on the low speed PIV databases are detailed in Table IV.5 and Table IV.6. It is difficult to clearly establish a relation between the upstream velocity and the determination coefficients from these data. Globally, R^2 seems to increase when the upstream velocity decreases. Values obtained for U_0 = 10 or 15 m.s^{-1} are higher than values at U_0 = 30 m.s^{-1}. But the evolution is not continuous, the lowest values (for the vertical plane) are those at U_0 = 20 m.s^{-1}. These discrepancies probably come from the uncertainties due to the lack of convergence of the data that was observed in Chapter II.3.3.4.1.

Upstream velocity (m.s^{-1})		10	15	20	25	30
Reynolds number (Re_h)		2.02×10^4	3.04×10^4	4.05×10^4	5.06×10^4	6.08×10^4
R^2 from original fields	Training	16.1	16.7	13.6	14.8	14.7
	Validation	15.7	15.8	12.4	13.9	14.5

Table IV.5: Determination coefficient evolution as function of the upstream velocity (LSE, low speed PIV database, vertical plane)

Upstream velocity (m.s^{-1})		10	15	20	25	30
Reynolds number (Re_h)		2.02×10^4	3.04×10^4	4.05×10^4	5.06×10^4	6.08×10^4
R^2 from original fields	Training	21.1	21.4	18.1	16.8	16.7
	Validation	19.7	21.2	16.5	17.2	15.7

Table IV.6: Determination coefficient evolution as function of the upstream velocity (LSE, low speed PIV database, horizontal plane)

R^2 values obtained using LSE on the high speed PIV databases are displayed in Table IV.7 and Table IV.8. Only three upstream velocities are tested and the results seem more consistent concerning the evolution of R^2 with the upstream velocity. When the upstream velocity decreases, R^2 increases. Also, R^2 calculated on the validation set is for some cases higher than the value from the training set. That comes from the fact that the validation set is smaller than the training set. These values are probably not converged and one should not expect to obtain a better accuracy on the validation set than on the training set in general. Nevertheless, the global evolution of R^2 as function of U_0 remains the same on both sets. At last, the discrepancies between R^2 values obtained from the low speed PIV database and the high speed PIV database also comes from the fact that the velocity fields estimated do not have the same size and resolution in both databases. The fact that the resolution is higher with the high speed PIV leads to spatially filtering out some small scale structures that were contained in the low speed PIV. Thus, it is expected to increase R^2. In addition, high speed PIV snapshots are shorter in the streamwise direction and the error is globally higher far downstream the cube than close to it. This also should lead to obtain higher R^2 using the high speed PIV

data instead of the low speed ones. That is observed by comparing R^2 value between low speed PIV and high speed PIV cases (especially for the horizontal plane).

Determination coefficients of the QSE are shown in Table IV.9 and Table IV.10 for the high speed PIV databases. As expected, training set R^2 is increased compared to the LSE case. Validation set R^2 is also increased but the values have to be considered with care as it was just explained. Nevertheless, the same evolution as for the LSE is observed. The estimation becomes more accurate when the Reynolds number is decreasing. These observations are expected. Indeed, the Kolmogorov scales decreases when the Reynolds number increases, while the largest remains approximately the same (they are imposed by the cube geometry). Thus energy is shared over a larger range of scales when the Reynolds number increases. Since it was shown that the shortest length scales are estimated with a lower accuracy than the largest ones by SE methods, the R^2 decreases. The large scales that are well estimated represents a smaller part of the flow energy content.

Upstream velocity ($m.s^{-1}$)		10	20	30
Reynolds number (Re_h)		2.02×10^4	4.05×10^4	6.08×10^4
R^2 from original fields	Training	17.5	16.3	14.7
	Validation	20.4	17.6	13.4

Table IV.7: Determination coefficient evolution as function of the upstream velocity (LSE, high speed PIV database, vertical plane)

Upstream velocity ($m.s^{-1}$)		10	20	30
Reynolds number (Re_h)		2.02×10^4	4.05×10^4	6.08×10^4
R^2 from original fields	Training	33.5	23.8	21.8
	Validation	33.4	24.3	23.5

Table IV.8: Determination coefficient evolution as function of the upstream velocity (LSE, high speed PIV database, horizontal plane)

Upstream velocity ($m.s^{-1}$)		10	20	30
Reynolds number (Re_h)		2.02×10^4	4.05×10^4	6.08×10^4
R^2 from original fields	Training	21.6	20	18.5
	Validation	23.3	19.5	14.6

Table IV.9: Determination coefficient evolution as function of the upstream velocity (QSE, high speed PIV database, vertical plane)

Upstream velocity ($m.s^{-1}$)		10	20	30
Reynolds number (Re_h)		2.02×10^4	4.05×10^4	6.08×10^4
R^2 from original fields	Training	38.7	28.5	26.5
	Validation	36.6	27.2	24.2

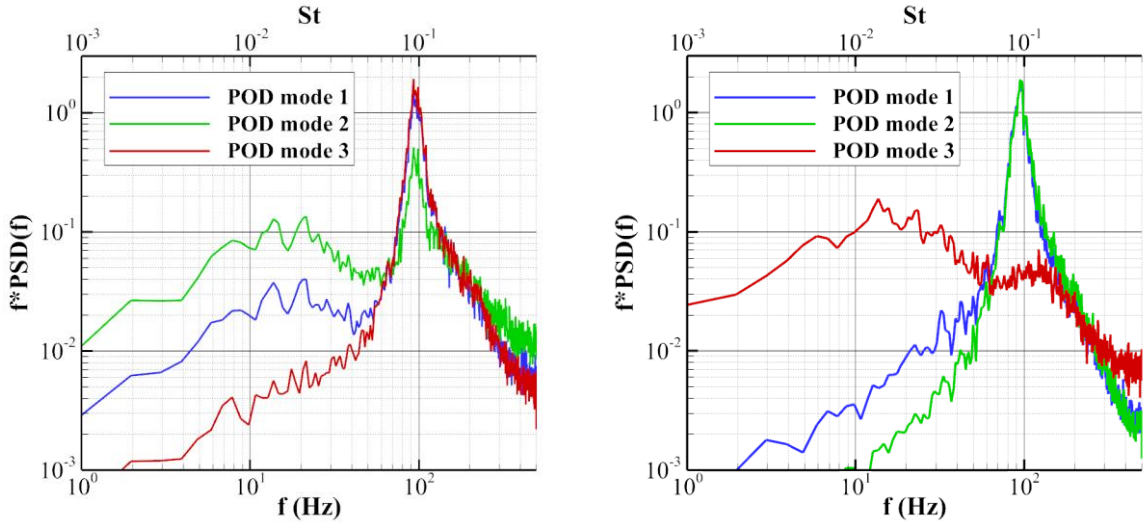
Table IV.10: Determination coefficient evolution as function of the upstream velocity (QSE, high speed PIV database, horizontal plane)

5.2. Flow characterization using Proper-Orthogonal-Decomposition

Estimation, using modified LSE (LSE-POD), of the velocity field around a wall mounted cube was already investigated in Chapter III.4.2. It was shown that for both planes the first three POD modes were estimated with a determination coefficient much higher than for POD modes of

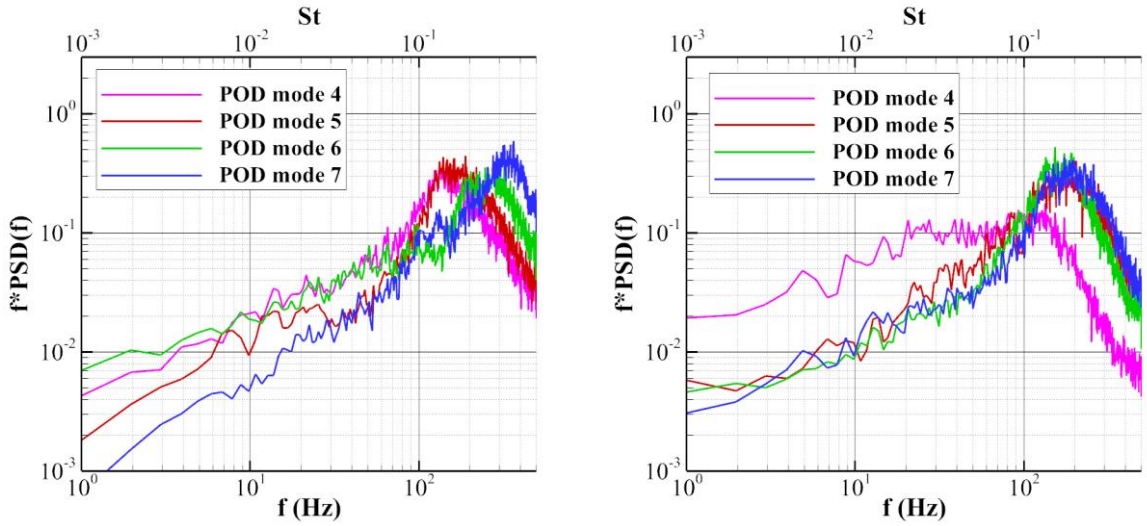
higher ranks. The PSD of these POD modes are plotted in Fig. IV.57. The PSD of POD modes 4 to 7 are plotted in Fig. IV.58. All the spectra are already normalized since the norm of each POD coefficients is one. For the vertical plane, the first three POD modes exhibit a clear peak at f = 100 Hz (St = 0.1) which is the shedding frequency. POD modes 1 and 2 have also more energy in the low frequencies (about 20 Hz). PSD of POD modes 4 to 7 have no peak at the shedding frequency and the spectrum of these POD modes is clearly translated towards higher frequencies. A bump is observed around 200-400 Hz (depending on the POD modes). Therefore, as in the OAT15A airfoil test case, the best estimated POD modes are those containing the lowest frequencies of the flow. That is also observed for the estimation of the horizontal plane. Among POD modes 4 to 7, for the vertical plane, the POD mode 7 is predicted with the lowest accuracy. Its PSD shows that its frequency content is at higher frequencies than the ones of POD modes 4, 5 or 6. In the BFS case, no POD modes are predicted with R^2 higher than 25% as it is the case for the first three POD modes in the cube experiments. It seems that it may be related to the frequency content of the modes. In the BFS case, POD modes have a quite broad spectrum (see Fig. H.5 in annex H.3) contrary to what is observed for the first three POD modes in the cube test case.

Differences in the prediction accuracy between the first three POD modes of both planes are more difficult to explain. Indeed, for the horizontal plane, POD mode 3 is the best predicted. Its PSD is quite different than PSD of POD modes 1 and 2. No peak at the shedding frequency is observed but there is a strong bump around 20 Hz. Thus, POD mode 3 contains more low frequencies (under 100 Hz) than POD modes 1 and 2. However, for the vertical plane, POD mode 2 is predicted with the lowest R^2 while it contains more low frequencies than POD modes 1 and 3. But this mode also has a clear peak at the shedding frequency, which is not the case for POD mode 3 of the horizontal plane.



a) Vertical plane b) Horizontal plane

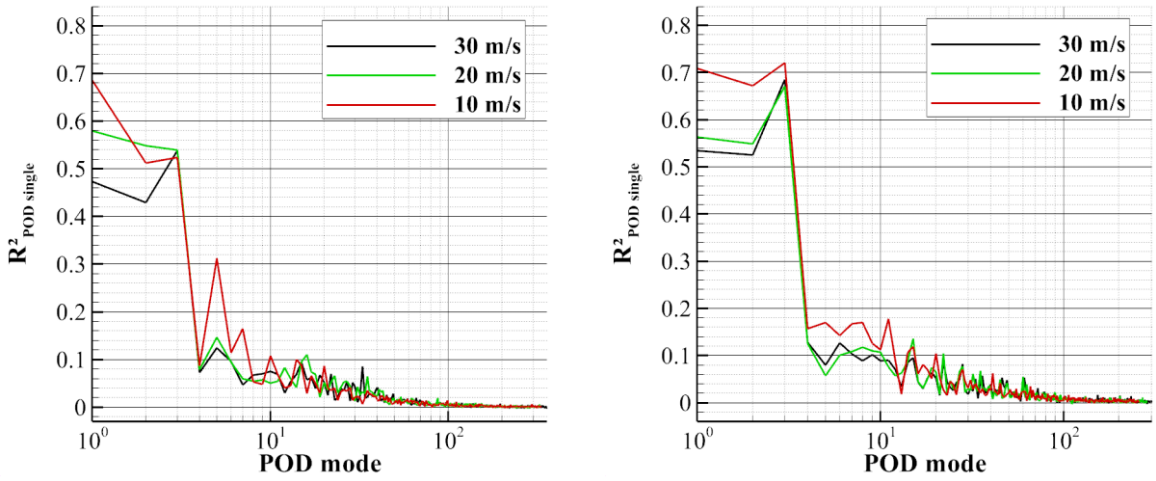
Fig. IV.57: Power Spectral Density of POD modes 1 to 3 (High speed PIV database, U_0 = 30 m.s^{-1})



a) Vertical plane b) Horizontal plane

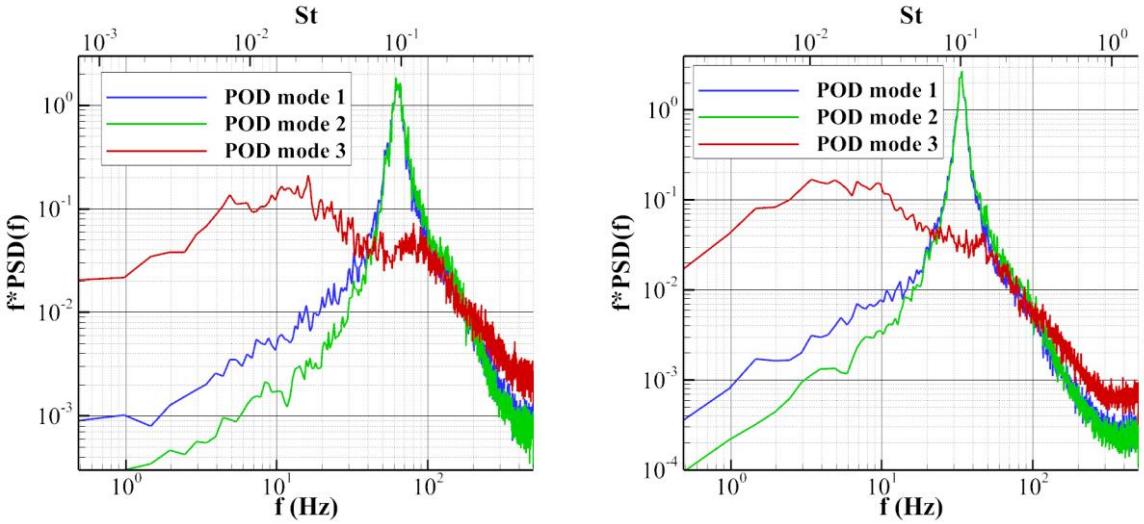
Fig. IV.58: Power Spectral Density of POD modes 4 to 7 (High speed PIV database, U_0 = 30 m.s^{-1})

Determination coefficients calculated for POD modes separately are plotted in Fig. IV.59. In this figure, results for the three upstream velocity conditions are shown. For all velocities, R^2 evolution as function of the POD mode remains globally unchanged. The first three POD modes are always the best reconstructed and predicted. For POD modes of higher ranks, there is a clear deterioration of the estimation accuracy. And the estimation of POD modes of ranks higher than 70 is completely inaccurate (R^2 is close to 0). Looking at the prediction accuracy of the first three POD modes of the horizontal plane, the order, from best predicted to worst predicted POD mode, is unchanged between the three situations. The PSD of these POD modes are displayed in Fig. IV.60 for the 10 and 20 m.s^{-1} case. This figure shows that the PSD of these three POD modes are perfectly identical between the three situations (when the Strouhal number is used as abscissa). The only difference is that spectra are translated toward lower frequency when the upstream velocity decreases. On the contrary, for the vertical plane, the order, from the best predicted to the worst predicted POD mode, is changed between the three tested situations. For instance, the best predicted POD mode is the third one in the 30 m.s^{-1}, case but it is the first one in the two other cases. The PSD of the first three POD modes of the vertical plane are plotted in Fig. IV.61 for the 10 and 20 m.s^{-1} case. Those spectra show that the best predicted POD mode always corresponds to one with the lowest levels around St ≈ 0.02 (see Fig. IV.57 for the 30 m.s^{-1} case). On the contrary, the worst predicted mode always has the highest levels around St ≈ 0.02. These observations induce a clear relation between the frequency content of a POD mode and the accuracy of its prediction. Between these three POD modes, the best predicted is therefore not the one that contain more energy in the low frequency than the other two. It seems here that POD modes with broad spectra are estimated with lower accuracy than POD modes with narrow spectra. However, that does not contradict previously stated conclusions about the difficulty to estimate the high frequency content. Indeed, the first three POD modes that mainly contains frequencies lower than the main phenomenon (St ≈ 0.1) are clearly the best estimated. The other POD modes contains higher frequencies.



a) Vertical plane b) Horizontal plane

Fig. IV.59: Determination coefficient calculated for each POD mode separately on the validation set using LSE for several upstream conditions (wall mounted cube experiment, high speed PIV database)

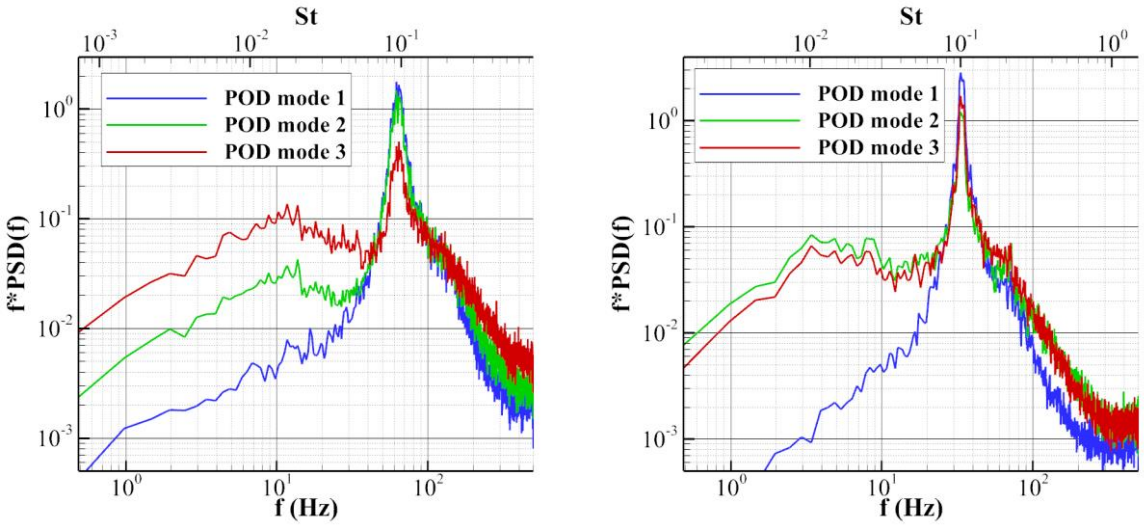


a) U_0 = 20 m.s^{-1} b) U_0 = 30 m.s^{-1}

Fig. IV.60: Power Spectral Density of POD modes 1 to 3 (High speed PIV database, horizontal plane)

Some improvements, in terms of R^2, are also observed between the three situations, especially between the 10 m.s^{-1} case and the two others. From Fig. IV.59, it is clear that some POD modes are better predicted in the 10 m.s^{-1} case than in the 20 or 30 m.s^{-1} cases. That is consistent with the evolution of R^2 computed from the entire velocity field displayed in Table IV.7 and Table IV.8. However, improvements mainly concern POD modes of ranks lower than 10 and more particularly the first two POD modes. The gains are not so directly linked to the spectra of POD modes. The fact that when the upstream velocity decreases, the spectrum of a POD modes is generally translated toward the low frequency may explain for a part the increase of R^2. But that is not sufficient to completely explain some observations. For instance, while the evolutions of POD modes 4 and 5

spectra are similar, for the vertical plane, between the 10 m.s^{-1} and the 30 m.s^{-1} cases, only the reconstruction and prediction accuracy of the POD mode 5 is improved.



a) U_0 = 20 m.s^{-1} b) U_0 = 10 m.s^{-1}

Fig. IV.61: Power Spectral Density of POD modes 1 to 3 (High speed PIV database, vertical plane)

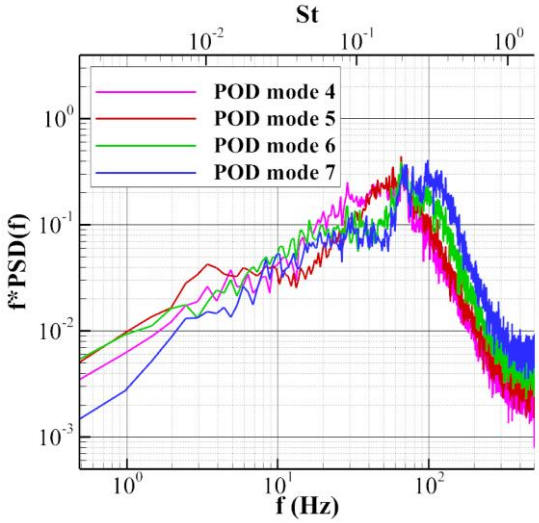
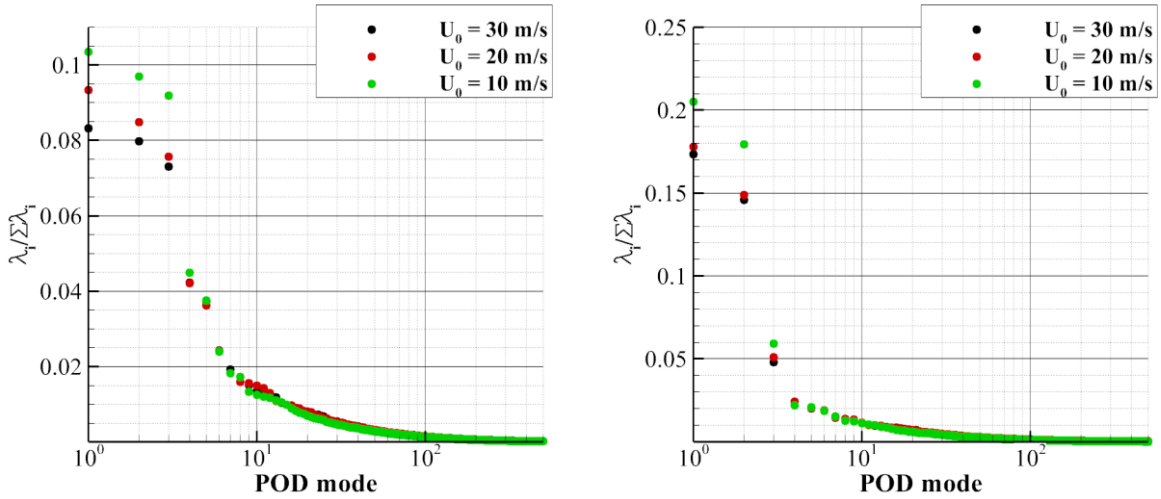


Fig. IV.62: Power Spectral Density of POD modes 4 to 7 (High speed PIV database, vertical plane, U_0 = 10 m.s^{-1})

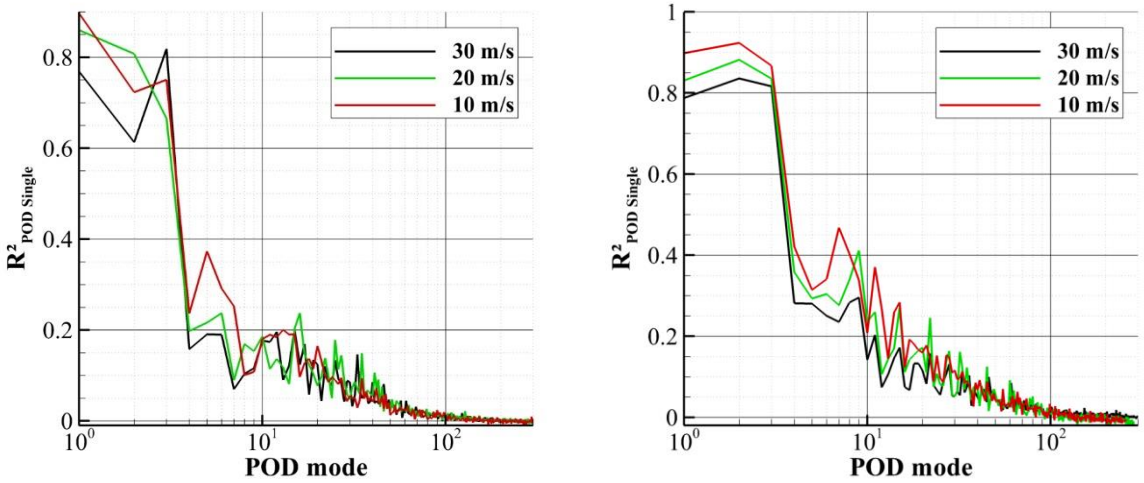
At last, another phenomenon acts for the improvement of accuracy of the estimation when the upstream velocity is decreasing. In addition to the increase of R^2 for the first POD modes, less and less POD modes are required to retain the same amount of TKE when the upstream velocity decreases. That is expected from the fact that the range of turbulent scales increases with the Reynolds number. Percent of TKE contained in each of the first 500 POD modes for both planes and the three upstream velocity cases are plotted in Fig. IV.63. It is clear from this figure that the first three POD modes of the vertical plane contain more and more energy when U_0 decreases. They contain 23.6%, 25.4% and 29.2% of TKE in the 30, 20 and 10 m.s^{-1} cases respectively. The same can be said for the first two POD modes of the horizontal plane (31.9%, 32.7% and 38.4% of TKE in the 30, 20 and 10 m.s^{-1} cases respectively). Therefore, even if the estimation accuracy of these POD

modes were not improved, the overall estimation will likely be more accurate when U_0 is decreased. Indeed, a larger part of the TKE will be estimated with high R^2. The accuracy improvement is thus the results of the fact that the first POD modes, that are the best predicted, contain more and more energy when U_0 decreases and also that they are themselves better predicted when U_0 decreases.



a) Vertical plane b) Horizontal plane

Fig. IV.63: Energy spectrum of the first 500 POD modes (Cube, high speed PIV)



a) Vertical plane b) Horizontal plane

Fig. IV.64: Determination coefficient calculated for each POD mode separately on the validation set using MTD-LSE for several upstream conditions (wall mounted cube experiment, high speed PIV database)

Determination coefficients of POD modes obtained using MTD-LSE-POD are plotted in Fig. IV.64. This figure can be compared with Fig. IV.59. From these figures, but also from Fig. III.79 and Fig. III.80, it appears that the use of delays improves mainly the first three POD modes for the vertical plane and the first ten POD modes for the horizontal plane, whatever the upstream velocity. The impact of using delays is perfectly similar for the three upstream velocity conditions tested.

5.3. Generalization of the SE to several upstream velocity conditions

In the previous sections, the direct impact of the upstream velocity on the prediction quality has been evaluated in the case where SE was used to predict the velocity field at the upstream conditions used to train it. We now look at the impact of using the SE, trained for some upstream velocity, to predict the flow at another upstream velocity. For brevity, only the horizontal plane data of the low speed PIV database are used. First the LSE is trained using the training data set at one upstream velocity. It is important to notice that the training data are not normalized so far. Then, using the LSE coefficients obtained, data from validation set at the other upstream velocity are predicted. The results are presented in Table IV.11. As one could expect, the prediction accuracy of the LSE decreases when one is estimating a velocity field which upstream conditions are more and more different to the ones used for the training. With such training, LSE should be used to predict only the same flow configuration.

Training U_∞ / Prediction U_∞	10	15	20	25	30
10	**19.7**	16.2	14	12.2	10.7
15	15.1	**21.2**	19.8	17.5	15.2
20	0.82	14.2	**16.5**	15.8	14.2
25	-14.9	-4.5	15.7	**17.2**	16.6
30	-39.2	-4	9.6	14.6	**15.7**

Table IV.11: Determination coefficients of several LSE trained at different upstream velocities using non normalized data (validation set)

The possibility to extend the use of a same set of LSE coefficients to several upstream velocities by training the LSE with normalized data is now assessed. To do so, the training data and validation data are normalized. The three components velocity fluctuations are normalized using the freestream velocity. The pressure fluctuations of the sensors are normalized by the freestream dynamic pressure. The determination coefficients obtained for the prediction of the normalized validation data of the several upstream velocity situations are shown in Table IV.12. The results shown in this table demonstrate that, when trained with normalized data, the LSE prediction accuracy outside the training set, for some other upstream velocity conditions, are less deteriorated than if non normalized data are used for the training. As such, LSE trained with the normalized 30 $m.s^{-1}$ dataset gives results very close to those obtained when training and validation sets are at the same velocity.

At last, LSE is trained using the ensemble of normalized training dataset at 10, 15, 20, 25 and 30 $m.s^{-1}$. The prediction accuracy of this LSE at these 5 velocities is shown in the column "LSE Mix" of Table VI.14. R^2 obtained with such a training are the same as the ones obtained when LSE is train for each velocity alone (except for the 10 $m.s^{-1}$ case). This way of training the LSE seems to be the best in order to be able to use it for the larger range of upstream velocity conditions. However, the results obtained when the upstream velocity is 10 $m.s^{-1}$ shows that LSE training for some conditions cannot be used for completely different conditions without deteriorating its prediction performances.

Training U_∞ / Prediction U_∞	10	15	20	25	30	"LSE Mix"	"QSE Mix"
10	**19.7**	14.9	13	13.3	14.6	14.6	18
15	17.9	**21.3**	20.9	20.8	20.6	21.2	23.7
20	12.7	16.2	**16.5**	16.5	16.2	16.5	18.3
25	13.6	17	17.2	**17.2**	17.1	17.2	19.4
30	12.7	15.4	15.5	15.6	**15.7**	15.7	18

Table IV.12: Determination coefficients of several LSE or QSE trained at different upstream velocities using normalized data (validation set)

QSE is now trained using the ensemble of normalized training dataset at 10, 15, 20, 25 and 30 $m.s^{-1}$. Results are displayed in the last column of Table VI.14. These results should be compared with those in Table III.18 and Table H.3, Table H.5, Table H.7, Table H.9 of annex H.4.1. This comparison shows that they are better than those obtained with a training using the data of the corresponding velocity only (even with Tikhonov). Therefore, this way of training the QSE allows to reduce the overfitting of the QSE and to obtained better results than with LSE for the 5 tested velocities. In this situation, it is clearly possible to use the same QSE coefficients for different velocity. In addition, it actually improves the prediction accuracy of the QSE. Nevertheless, these results remain lower than those obtained with MTD-LSE (trained for each upstream velocity separately).

Concerning the case of MTD-LSE, the training is more specific to the upstream velocity because the optimal number of delays to use depends on the main phenomenon period which itself depends on the upstream velocity. Therefore, it does not seem possible to optimize the MTD-LSE (in term of number of delays, time window and Tikhonov parameters) for several upstream velocities. Using MTD-LSE optimized for a precise upstream velocity, to estimate the flow at a different upstream velocity, will obviously result in lower accuracy than if the optimal MTD-LSE, at this second upstream velocity, was used.

6. Phase-averaged estimation

In their work, Ruiz et al. [59] show that phase averaged fields of their flow were very well estimated using what they called Spatio-Temporal-LSE (ST-LSE). To determine the phase at each instant, they used a POD decomposition of a set of pressure sensors enriched of delays in the past and the future called Spatio-Temporal-POD (ST-POD). The ST-POD coefficients are also used to perform the LSE which forms the ST-LSE method. Details on the method are given in annex E. In particular, it is demonstrated that if all ST-POD modes are considered then the ST-LSE is equivalent to the MTD-LSE.

In this part, the estimation of phase averaged fields by LSE is quickly investigated. In addition, we look at the opportunity to improve the LSE by using a different set of LSE coefficients for each phase interval. The high speed PIV database of the wall mounted cube is used. The upstream velocity considered is 10 $m.s^{-1}$ so that the number of snapshots by period of the main phenomenon is the largest. The high speed PIV database is preferred to the low speed one because more snapshots are then available by phase interval. Only the vertical plane case is considered for brevity but the results are expected to be similar for the horizontal plane.

6.1. Estimation of phase-averaged fields by LSE

First the ST-POD of the 23 pressure sensors is computed (on the training dataset). 31 delays in the past only, spaced in time every 1×10^{-3} s, are used and they span one vortex shedding period. As expected from the works of Perrin et al. [97] and Ruiz et al. [59], the first two ST-POD coefficients carry the same level of energy (approximately 12.7% and 12.2% respectively) and they are shifted by a quarter of period. Thus, they can be used to determine the phase of the flow using the following formula:

$$\varphi_{a_1-a_2} = \arctan\left(\frac{a_2}{a_1}\right) \tag{IV.5}$$

These two ST-POD coefficients are displayed in Fig. IV.65 as well as the phase between them. Using this phase, the PIV snapshots can then be classified according to the phase obtained from the synchronized pressure measurements. The same can be done for the estimated snapshots (even outside the training set by projecting the pressure measurements on the ST-POD modes). The snapshots in the same ensemble are then averaged to form the phase average fields. Six phase intervals of same length are considered $\left(\left[\frac{2\pi}{6}i, \frac{2\pi}{6}(i+1)\right[, i \in [\![1,6]\!]\right)$. The phase averaged field of the first interval is plotted in Fig. IV.66. The phase averaged of the LSE estimated fields from the validation set for the same phase interval is shown in Fig. IV.67. Only the vertical velocity is displayed. Qualitatively, the phase averaged field seems to be well predicted by the LSE. The main structures are predicted, the level of fluctuations of the streamwise velocity are however smaller.

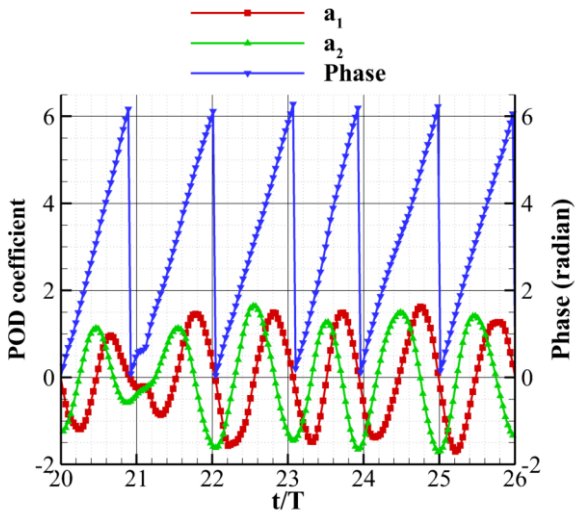


Fig. IV.65: Evolution of the first two ST-POD coefficients and of the phase between them

Determination coefficients of the prediction by LSE of the phase averaged fields of the six intervals are given in Table IV.13. They confirm the good prediction of the phase average fields. These results are in agreement with the observations of Ruiz et al. and demonstrate that the phase information can be obtained with LSE with a good accuracy even if each instantaneous velocity snapshots is not well predicted. Indeed, while R^2 is 20.4% (see Table H.10 in annex H.4.2) for the instantaneous velocity field, it is around 85% for the phase averaged fields.

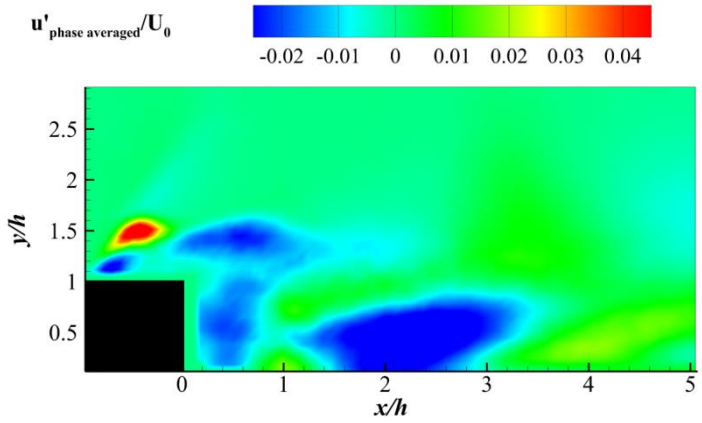


Fig. IV.66: Phase averaged field (from original training data set) for the interval $\left[0,\frac{2\pi}{6}\right[$ (streamwise velocity component)

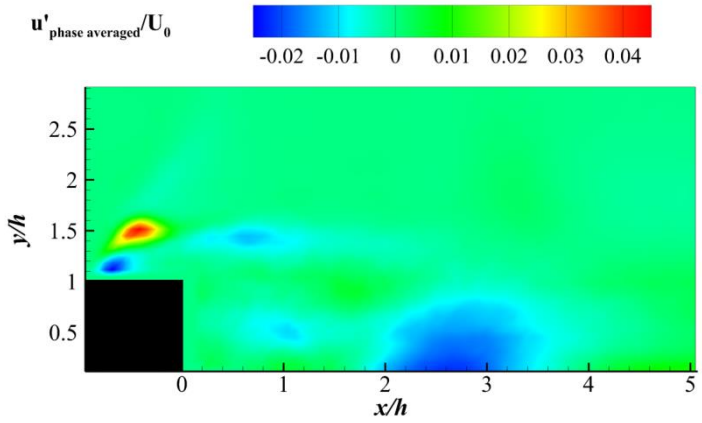


Fig. IV.67: Phase averaged field (from LSE estimation of the validation set) for the interval $\left[0,\frac{2\pi}{6}\right[$ (streamwise velocity component)

	$\left[0,\frac{\pi}{3}\right[$	$\left[\frac{\pi}{3},\frac{2\pi}{3}\right[$	$\left[\frac{2\pi}{3},\pi\right[$	$\left[\pi,\frac{4\pi}{3}\right[$	$\left[\frac{4\pi}{3},\frac{5\pi}{3}\right[$	$\left[\frac{5\pi}{3},2\pi\right[$
R^2 (%)	84.4	88.2	88.3	86.7	87.8	88.8

Table IV.13: Determination coefficients of the prediction by LSE of the phase averaged fields

6.2. Improving the LSE using phase information

In this part, a small modification of the LSE is proposed in the hope of improving the LSE accuracy. It has been shown that the first two coefficients of the ST-POD can be used to determine in real time a phase for the flow. The idea is thus to use different sets of LSE coefficients for different phase. To do so, a number of phase intervals is chosen. Then the data in the training set are split into the same number of sub training sets according to their phase. At last, the LSE is trained for each subset (corresponding to one phase interval) separately, thus forming the same number of LSE coefficient sets. To estimate a velocity field, its phase is thus first calculated using the projection onto the first two ST-POD modes of the pressure measurements, then the field is estimated by LSE using the set of LSE coefficients corresponding to the calculated phase. We called this method "phased"-LSE. Concerning the convergence of the data to train the "phased"-LSE, it should not be worse than for the classic LSE when only one set of LSE coefficients is considered. Indeed, there

is only one over 30 samples that are statistically independent, thus as long as the number of phase intervals considered is less than 30, the number of statistically independent samples in each subset should be of the same order of magnitude than the one in the initial training set.

Six situations are investigated in which the phase is split into 2, 6, 10, 20, 30 and 40 intervals. The prediction accuracy in term of determination coefficient of the velocity fields are summarized in Table IV.14. In the best case (10 and 20 intervals), R^2 is increased by about 4% compared to the LSE using only one set of coefficients. It is therefore possible to improve the accuracy of the LSE by using the "phased"-LSE. Nevertheless, in the present test case, the improvement is limited and it is unclear how much better the results could have been if data sampled at a higher frequency were available. Also, at first when increasing the number of phase interval R^2 increases. This shows the interest of using LSE coefficients that are specialized for a reduced phase interval. However, when the number of phase interval becomes too large (especially larger than the number of samples by period of the lowest frequency phenomenon), R^2 decreases. It is probably due to the overfitting that becomes more important since the number of statistically independent samples by training subsets decreases. The "phased"-LSE results are also more accurate than the QSE ones but not than the MTD-LSE ones.

Number of phase intervals	2	6	10	20	30	40
R^2 (%)	21.6	24.2	24.6	24.6	24.3	23.8

Table IV.14: Determination coefficients of the prediction of instantaneous velocity fields by "phased"-LSE

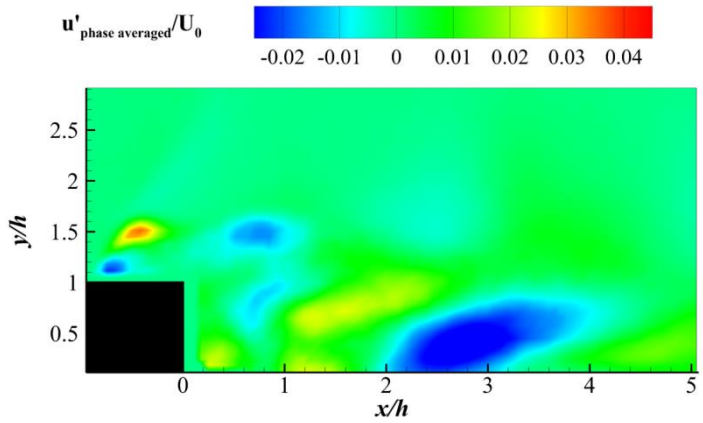


Fig. IV.68: Phase averaged field (from "phased"-LSE estimation of the validation set) for the interval $\left[0, \frac{2\pi}{6}\right[$ (streamwise velocity component)

	$\left[0, \frac{\pi}{3}\right[$	$\left[\frac{\pi}{3}, \frac{2\pi}{3}\right[$	$\left[\frac{2\pi}{3}, \pi\right[$	$\left[\pi, \frac{4\pi}{3}\right[$	$\left[\frac{4\pi}{3}, \frac{5\pi}{3}\right[$	$\left[\frac{5\pi}{3}, 2\pi\right[$
R^2 (%)	96.7	96.4	97.2	96.1	95.8	96.8

Table IV.15: Determination coefficients of the prediction by "phased"-LSE of the phase averaged fields

At last, the phase averaged fields obtained by phase averaging the "phased"-LSE estimated instantaneous fields are once more compared to the original phase averaged fields. The phase averaged field obtained by the "phased"-LSE when using 6 intervals is plotted in Fig. IV.68 for the first interval and the streamwise velocity. Determination coefficients in the 6 intervals case are given in Table IV.15. This figure and even more the determination coefficient values show that the prediction of the phase averaged fields is strongly improved using the "phased"-LSE. Moreover, it shows

that, while the prediction of the instantaneous velocity fields is not particularly accurate, it however contains a large part of the phase averaged information.

7. Chapter summary

The impact of the presence of a Gaussian noise on the data (both conditional events and data to estimate) has been investigated. The Gaussian noise contained in the data to estimate used to train the method is filtered out. Thus, if the noisy data are used to evaluate the performances of the SE then the performances will be underestimated. More importantly is the noise contained in the conditional events. It was shown that such noise directly penalizes the estimation accuracy. It is therefore important to use conditional events as clean as possible from noise. At last, one has to be very careful when using clean data, during the training, if the method is expected to be used in a noisy environment. That is typically the situation one encounters when training the SE with numerical data and then uses it on an experiment or real application. In this case, the noise contained in the conditional events outside the training set drastically deteriorates the estimation and the method appears to also suffer from overfitting. In such situation, the use of regularization can decrease the sensibility of the method to the noise in the conditional events.

Estimations of the flow around the OAT15A airfoil in transonic conditions obtained using URANS and ZDES simulations were then compared. The results showed that URANS fields were estimated with higher accuracy than ZDES fields. The study of the spectral content of the flow highlighted that the low frequency content was better reconstructed and predicted by the SE than the high frequency content (and its extensions). This fact explains why the URANS flow fields were better estimated than the ZDES ones. Indeed, ZDES simulated flow presents more energy in the high frequencies (particularly around the shedding frequency) than the URANS one. Also it was clearly observed that the resolution of part of the turbulence in the ZDES simulation leads to a shortening of the turbulent spatial integral length scales contained in the wake. The comparison of the integral length scale maps with maps of the normalized mean square error showed a relation between high error areas and short integral length scale areas. That indicates that short integral length scales are estimated with less accuracy than the long ones, and also participates to explain the difference in estimation quality between the URANS flow fields and the ZDES ones.

Following these observations, the conservation of the length scales by the LSE was more precisely investigated. It was shown that LSE does not conserve the turbulent spatial integral length scales and filters, in some way, the shortest lengths. These conclusions were also supported by studying the estimation of POD modes. A characteristic length can be associated to each POD mode (in the Backward Facing Step case) and it was observed that POD modes with the highest characteristic lengths were the best estimated. In addition, this study highlighted the strong impact of the sensor locations on the quality of the estimation using SE methods. As such, it was noticed that POD modes were estimated with higher accuracy when conditional events located at proximity of their spatial extrema were used. Therefore, it also explains the difficulty that rises when one tries to estimate the flow using only wall information. The optimization of the sensor locations will thus be addressed in the next chapter (Chapter V). At last, the study of the spatial filtering of the velocity fields also confirmed that the shortest structures of the flow are poorly estimated or even filtered out.

The temporal filtering of both conditional events and data to estimate was also investigated. Through this study, the low frequency content once more appears to be the best reconstructed and predicted. Filtering the conditional events (with a temporal low pass filter) was shown to slightly increase the accuracy as part of the noise was probably filtered out. However, filtering the data to estimate temporally or spatially did not lead to any significant improvements.

The impact of the Reynolds number on the estimation quality was investigated using data from the cube experiment. It was shown that the estimation became more accurate when the Reynolds number is decreased. When looking at the estimation of POD modes, the accuracy of their estimations improves with decreasing Reynolds number. In addition, the number of POD modes required to retain a same level of TKE decreases with Re (that is due to the fact that the flow contains less and less small scales). This also acts to improve the estimation accuracy as the POD modes, that are the best estimated, represents a larger and larger part of the flow when Re decreases. The possibility to use SE to estimate flow for which it does not have been trained was also studied. It was demonstrated that to some extent, training the LSE with normalized data allowed to predict the flow at different upstream conditions without a strong deterioration of the accuracy. Also, training the QSE using normalized data from several upstream velocity reduced the overfitting of the QSE model.

At last, the use of Spatio-Temporal POD to determine the phase of the velocity field in the wall mounted test case was studied and used to form phase-averaged fields. It was shown that the predicted instantaneous fields, while not particularly informative on the true instantaneous fields, still contains a large part of the information of the phase averaged fields. A modification of the LSE, called “phased”-LSE was also proposed and its accuracy was shown to be higher than the LSE and QSE one. The increase in R^2 was however limited and the “phased”-LSE results were less accurate than the MTD-LSE ones. It was not attempted to use multi-time-delays with the “phased”-LSE. One can fairly expect that the MTD-“phased”-LSE performances will be higher than the MTD-LSE ones. Also, the “phased”-LSE was able to retain even more information on the phase-averaged fields than the LSE.

Chapter V. Conditional event optimization

In chapter III and chapter IV, it was shown that the sensor locations have a strong impact on the reconstruction and prediction accuracy of SE methods. Therefore, in this chapter, the opportunity to use an optimization algorithm, based on multi-variate least square regression, to choose the sensor locations in LSE is investigated. An extension of the algorithm is then proposed to choose not only the sensor locations but also delayed information at the sensor locations.

Discussion on the algorithm, coupled with previous study of the length scale conservation, serve to propose a strategy to use modified LSE with or without delays.

1. Sensor location optimization

1.1. Sensor location optimization algorithm description

The algorithm studied in this chapter has been proposed by Muradore et al. [98] in the chemical field. The algorithm is a data driven one, based on least square regression. It is thus interesting for LSE applications since LSE can be seen as an ordinary least square regression.

The main idea of the algorithm is that, if data from several possible sensor locations are available, the algorithm will iteratively choose locations among the possible ones which have the highest correlation with the remaining part (in the sense that it has not yet been estimated by the previously chosen sensors) of the original data to estimate. Therefore, at each iteration, all the remaining possibilities have to be tested which could be computationally expensive.

To get a quick overview of the algorithm, let us assume that N time realizations of N_S sensors (X_1) are known, as well as N time realizations of a discrete scalar field (U_1), of spatial dimensions M. This scalar field is the data that has to be estimated. These data can be stored in the following matrices:

$$X_1 \in \mathbf{R}^{N_S \times N} \text{ and } U_1 \in \mathbf{R}^{M \times N}$$

In addition, it is assumed that mean has been removed to X_1 and U_1. Now, let us consider the i^{th} iteration, the algorithm first step is to compute the correlation matrix between the two previous data sets:

$$\Gamma_{X_i U_i} \in \mathbf{R}^{(N_S - i + 1) \times M}$$

If the data have been scaled so that all rows have unit variance, then $\Gamma_{X_i U_i} = \frac{1}{N} X_i U_i^T$, if not, then $\Gamma_{X_i U_i} = \Lambda_{X_i}^{-1/2} \left(\frac{1}{N} X_i U_i^T \right) \Lambda_{U_i}^{-1/2}$ where $\Lambda_{X_i} = \mathrm{diag}(\Sigma_{X_i})$ and $\Lambda_{U_i} = \mathrm{diag}(\Sigma_{U_i})$. Σ denotes the variance.

Once the correlation matrix computed, one sensor is chosen so that its correlation with the field is the maximum correlation. Thus the line number in X_i of the chosen sensor is:

$$j_i = \underset{1 \le j \le N_S - i + 1}{\arg\max} \sum_{k=1}^{M} \left| \left(\Gamma_{X_i U_i} \right)_{j,k} \right| \tag{V.1}$$

It is important to notice here that, if the scalar field is not spatially sampled such as all cell areas (or volumes) are equals, then the summation should be weighted using cell areas (or volumes).

Once the sensor chosen, both data matrices are updated. First the scalar field data are updated using an Ordinary Least Square (OLS) regression:

$$U_{i+1} = U_i - \widehat{U}_i = U_i - U_i X_i^{\mathrm{T}}(j_i) \left(X_i(j_i) X_i^{\mathrm{T}}(j_i) \right)^{-1} X_i(j_i) \tag{V.2}$$

where $X_i(j_i)$ is the line of X_i corresponding to the chosen sensors. $\widehat{U}_i$ is the OLS estimator (it is also the linear stochastic estimate of U_i using the sensor $X_i(j_i)$).

Then the sensor matrix is updated:

$$X_{i+1} = X_i^C - X_i^C X_i^{\mathrm{T}}(j_i) \left(X_i(j_i) X_i^{\mathrm{T}}(j_i) \right)^{-1} X_i(j_i) \tag{V.3}$$

where:

$$X_i^C = \begin{pmatrix} X_i(1) \\ \vdots \\ X_i(j_i - 1) \\ X_i(j_i + 1) \\ \vdots \\ X_i(N_S - i + 1) \end{pmatrix}$$

Updating the sensor matrix in such a way ensures that the sensor choice will be based on the information contained in the remaining sensors and not already contained in the previously chosen ones.

At last the algorithm is stopped if enough sensors are chosen, if the explained variance at iteration i is high enough, or if the explained variance variation is too weak (thus picking another sensor does not bring much more new information).

From the algorithm description, it appears that it is not directly linked to the flow physics and can be applied to any flow. The algorithm does not require any explicit knowledge about the dynamic of the flow. In addition, if the first chosen location corresponds to the most correlated sensor location with the data to estimate, the second choice is not the second best correlated sensor location. Indeed, the sensor data, as well as the data to estimate, are updated at each iteration. However, the selected sensors are each the most informative (in the least square sense) about the remaining part of the data to estimate from the remaining information in the sensors.

An important feature to highlight is that optimality cannot be formally guaranteed (at least not for data sets which have not been used to run the optimization sensor locations algorithm). Nevertheless, Muradore et al. concluded that, in the extensive simulations he showed, the algorithm picked the optimal situation.

1.2. Sensor location optimization algorithm applications

1.2.1. OAT15A airfoil in transonic conditions

The algorithm is applied to the choice of pressure sensors at the skin of the OAT15A airfoil. First, the algorithm is used on the data from the URANS simulation, then on the data from the ZDES. Also the algorithm is only used on the training set data.

- URANS simulation:

In this situation, the space of possible sensors is composed of the pressure information at the 644 mesh nodes forming the airfoil skin. The data to estimate is the velocity field containing both streamwise and vertical velocities. The algorithm is used to choose 5 pressure data in order to be compared with estimations made in Chapter III.2. In this chapter, two sets of 3 and 5 pressure sensors, which positions were chosen based on flow knowledge and intuition, were used.

The positions of the 5 chosen sensors are plotted in Fig. V.1. (Details on the positions are given in Table V.1, abscissae of the sensors are given (the origin being the leading edge of the airfoil), “s” denotes that the sensor is on the suction side and “p” denotes that it is located on the pressure side). The optimized sets of sensors are denoted “optim”. The order in which the sensors are chosen is also explained by coloring the sensors. The first sensor chosen is located on the pressure side at about half the chord. That is surprising but one has to remember that the algorithm is only based on the correlation. The correlation function in fact displays two peaks, one close to the nose of the airfoil and the other at the pressure side at half the chord. And the second one is slightly higher than the first one. Then, two sensors are chosen near the nose of the airfoil. At last, two sensors are picked downstream the shock on the suction side.

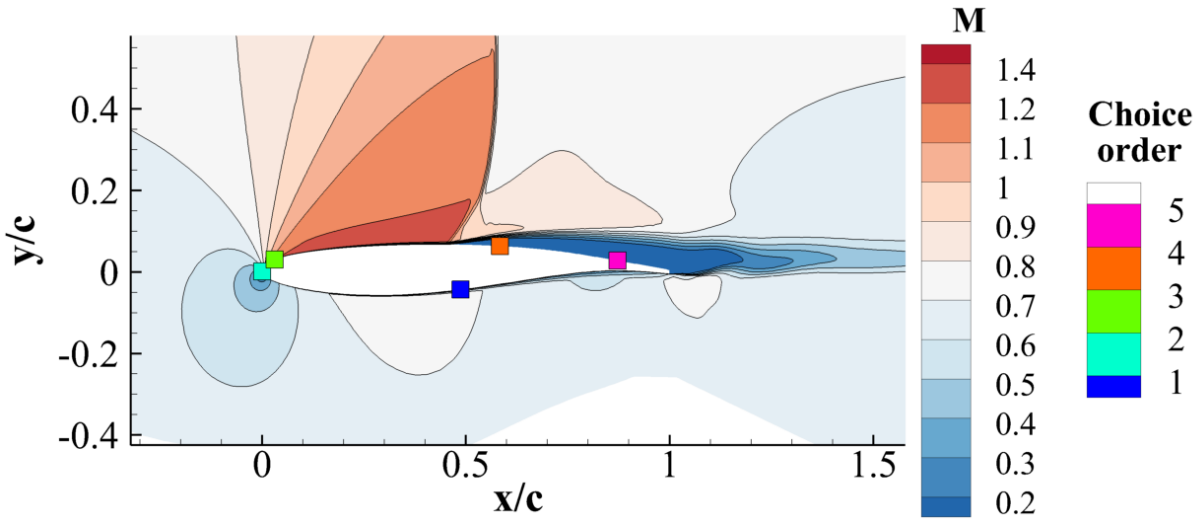


Fig. V.1: Positions of the 5 sensors chosen by the sensor location optimization algorithm in the URANS case (an instantaneous map of the Mach number is plotted in background with the shock in middle position)

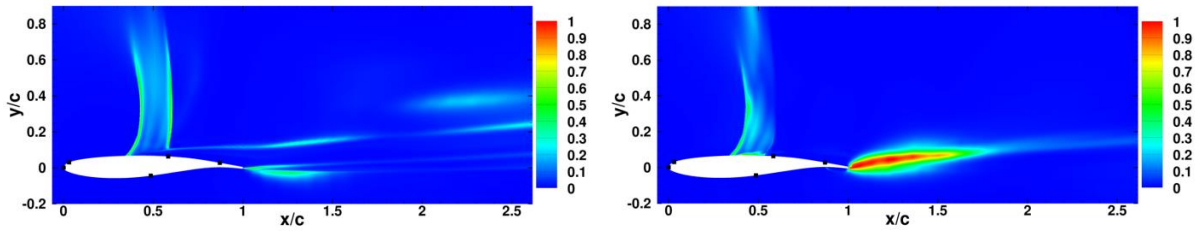
“Intuitive” P 5	“Optim” P 5
0.45 (s)	0.49 (p)
1 (s)	0 (s)
0 (s)	0.03 (s)
0.21 (s)	0.58 (s)
0.49 (p)	0.87 (s)

Table V.1: Sensors positions (in x/c) for the two sets of 5 pressure sensors (set of 3 pressure sensors are made of the first 3 sensors)

Determination coefficients obtained on the validation set using LSE and QSE are summarized in Table V.2 where the “intuitive” sets and the optimized sets of pressure sensors are used. These values show that LSE and QSE perform better when using the optimized sets of sensors than the intuitive ones. Improvements are more important for the LSE than for the QSE. Using LSE, R^2 is increased by 5% when 3 sensors are used and by 3% when 5 sensors are used. The validation set normalized mean square error of the LSE using the “optim” P 5 set is displayed in Fig. V.2. This figure should be compared with Fig. IV.8. From these two figures, it is clear that the prediction of the streamwise velocity is improved, especially in the wake and just downstream of the shock, when using the optimized set of sensors. However, improvements are mitigated for the vertical velocity. To complete the illustration, time evolutions of the streamwise velocity at point 1 (used in Chapter IV.2.1) and of the vertical velocity at point 2 using LSE and both sets of 5 pressure sensors are plotted in Fig. V.3. The optimized set clearly better predicts the streamwise velocity at point 1, but no real improvement is observed concerning the prediction of the vertical velocity at point 2. Both intuitive and optimized sets of sensors seem unable to predict the high frequency content of the flow. That can be seen on the PSD plotted in Fig. V.4.

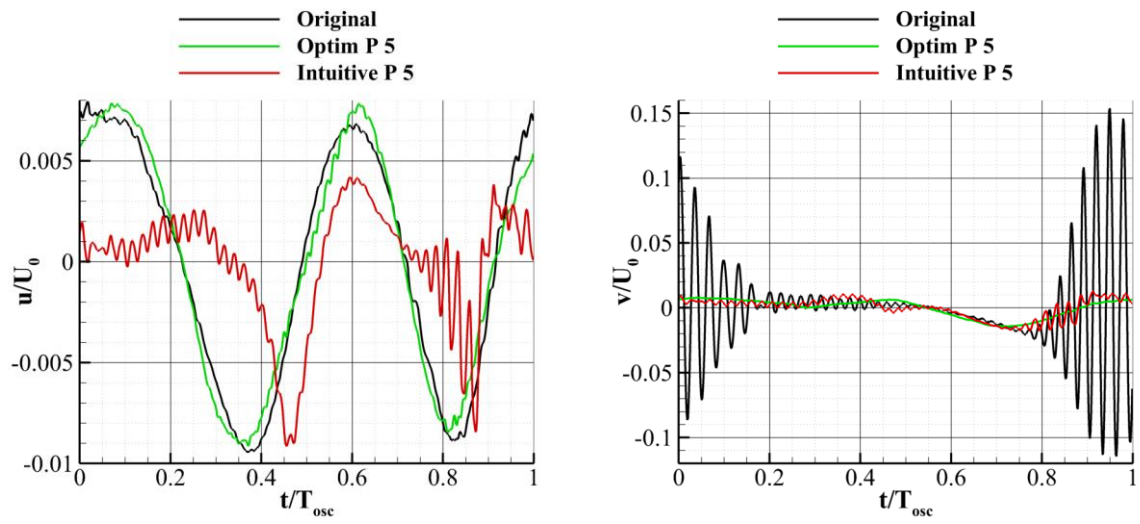
Sensors set	“Intuitive” P 3	“Optim” P 3	“Intuitive” P 5	“Optim” P 5
LSE	0.835	0.887	0.904	0.935
QSE	0.936	0.961	0.979	0.981

Table V.2: Determination coefficients R^2 comparison between several estimations using different sets of pressure sensors



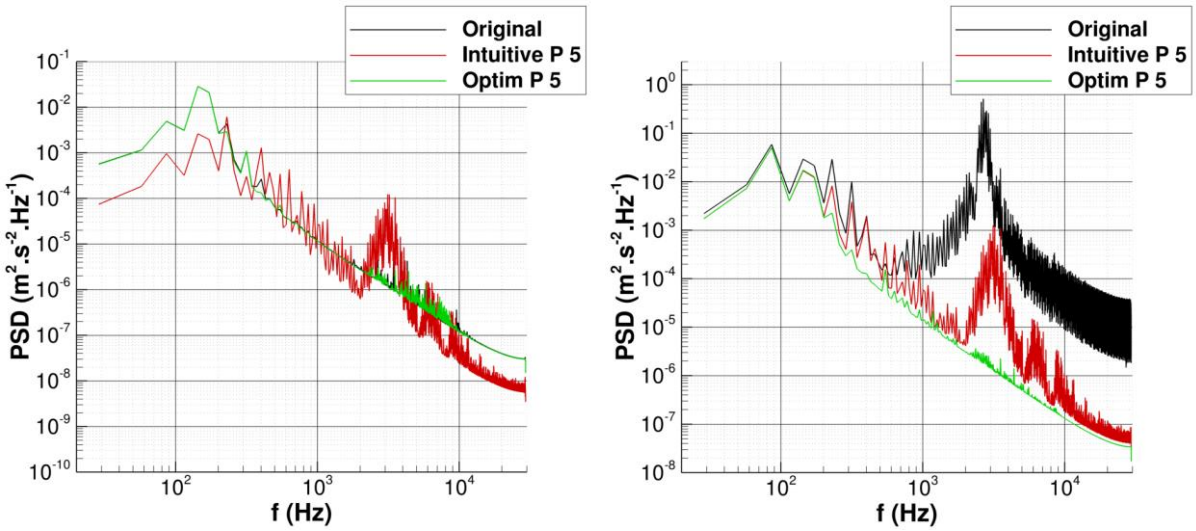
a) b)

Fig. V.2: Normalized mean square error of the LSE prediction of the URANS velocity field using the “optim” P 5 set of sensors. (a) streamwise velocity component, b) vertical velocity component)



a) Point (1, *u*) b) Point (2, *v*)

Fig. V.3: Streamwise and vertical fluctuating velocity signals comparison between original data and LSE predictions (URANS case)



a) Point (1, *u*) b) Point (2, *v*)

Fig. V.4: Power Spectral Density comparison between the original and the LSE predictions using the sets of sensors “Intuitive P 5” and “Optim P 5” (URANS case)

Values of R^2 calculated for the prediction of POD modes separately are shown in Fig. V.5. Using the optimized sets of pressure sensors leads to better prediction of POD modes 3 and 4 by LSE. Surprisingly the POD mode 6 R^2 is higher using the “intuitive” sets of sensors than the optimized ones. This observation could be explained by looking at the POD mode 6 and the sensor locations in Fig. V.6. As one can see in this figure, one sensor of the “intuitive” P 3 (and P 5) set is located nearby an extremum on the POD mode streamwise velocity. As previously explained, POD modes are better estimated when sensors are located near their extrema. However, looking at the shape of the POD modes and the sensor locations is not enough to explain the differences observed in Fig. V.5. For instance, no set of sensor has sensors particularly close to the extrema of the POD mode 4 but LSE using optimized sets is more accurate than LSE using “intuitive” sets.

When using MTD-LSE with the URANS simulated data, the impact of the sensor location appears less important due to the strong improvements brought by the addition of delays. As such, R^2 higher than 99% are obtained using the "intuitive" P 3 set as well as the "optim" P 3 set of pressure sensors.

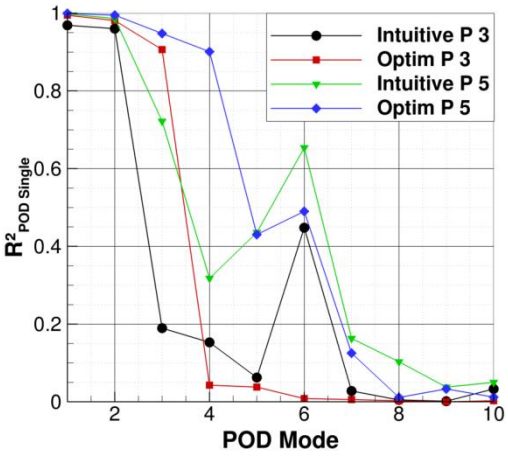
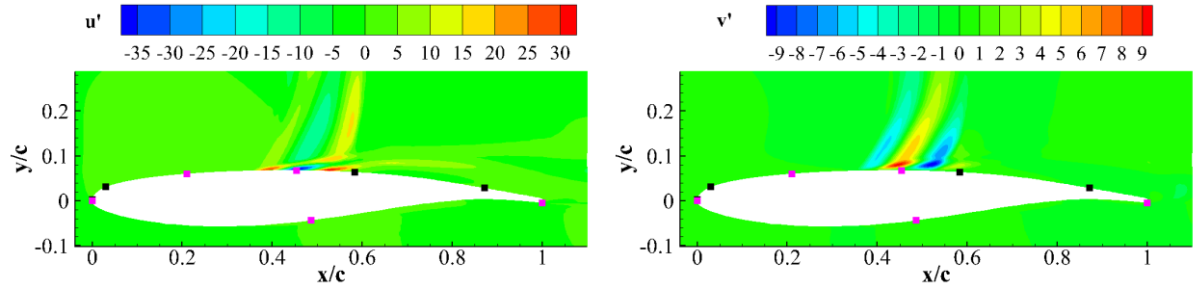


Fig. V.5: Determination coefficient calculated for each POD mode separately on the validation set using LSE (OAT15A URANS)



a) Streamwise velocity component b) Vertical velocity component

Fig. V.6: Contours of the streamwise (a) and vertical (b) velocity components of the POD mode 6. Pink squares correspond to "intuitive" P 5 set of sensors positions. Black squares correspond to "optim" P 5 set of sensors positions.

- ZDES:

Once more the space of possibilities for the sensors is made of the pressure data at all the mesh nodes of the airfoil wall. In total 411 locations are possible. The algorithm is used to optimize the estimation of the velocity field containing both streamwise and vertical velocities.

The first 5 locations chosen by the algorithm are shown in Fig. V.7, where map of the Mach number when the shock is at its most upstream position has also been plotted. Among those sensors none are chosen on the suction side of the airfoil, contrary to the URANS case. Two sensors (the second and the third) are chosen close to the noise, similarly to the URANS case. One is chosen at about half the chord on the pressure side. It is the first one and its position is very close to the one of the first chosen sensor in the URANS case. The fact that the first three sensors are chosen at similar positions when using ZDES data or URANS data is reassuring. Indeed, the physical phenomenon remains the same between the two simulations. The fourth and fifth sensors chosen in the ZDES case are not at all located around the same locations as in the URANS case. There are not on the same side. One thing to notice is that the two sensors chosen on the suction side in the URANS

case are located in the zone simulated using DES in the ZDES case (see Fig. II.58). The resolution of part of the turbulence by the DES may lead to a deterioration of the correlation and explain why sensors are chosen on the pressure side, where URANS equations are used.

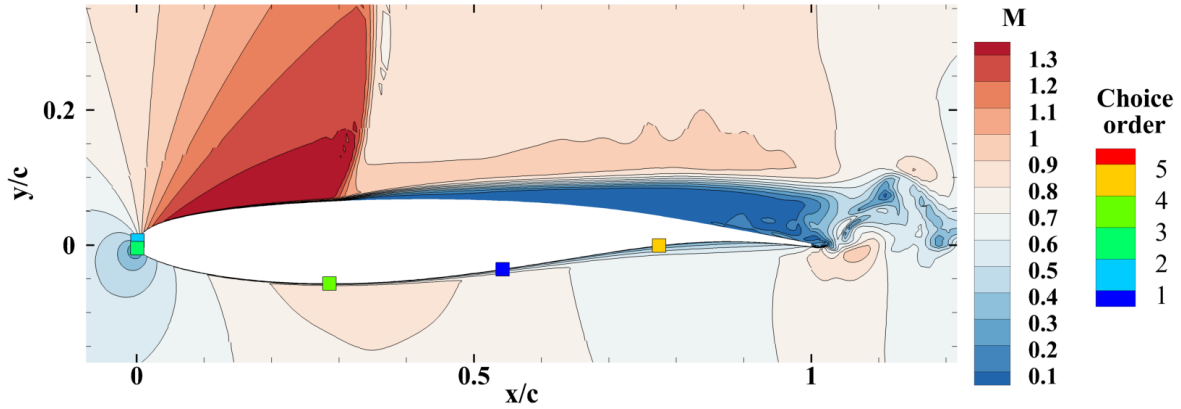
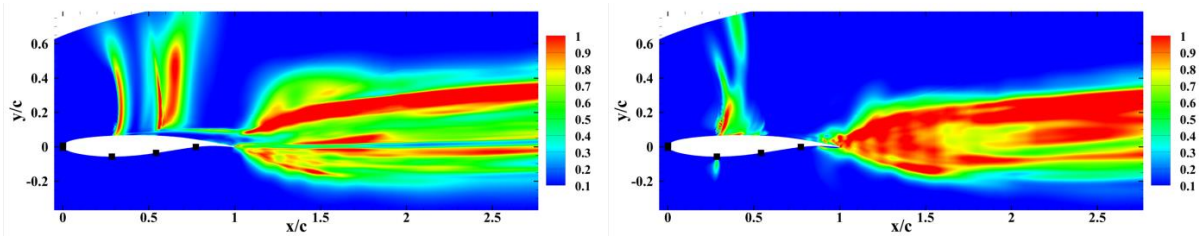


Fig. V.7: Positions of the 5 sensors chosen by the sensor location optimization algorithm in the ZDES case (an instantaneous map of the Mach number is plotted in background with the shock is at its most upstream position)

Validation set R^2 using LSE and QSE with the several sets of sensors are displayed in Table V.3. The highest R^2 are always obtained using the optimized sets when LSE is used. However, when QSE is used, the "intuitive" P 3 set performs better than the "Optim" P 3 one. This is not particularly surprising since the algorithm does not take into account second order relationship between the sensors and the field. The validation set normalized mean square error of the LSE using the "optim" P 5 set is displayed in Fig. V.8. This figure can be compared with Fig. III.16. The comparison of these two figures highlights that the error is decreased in the wake when the optimized set of sensors is utilized. On the contrary, the error around the shock is smaller when the "intuitive" set is used. This explains the small differences in determination coefficient between the two sets. Also, the locations obtained by the optimization algorithm do not ensure the best estimation in every point of the velocity field. That is not particularly surprising since the algorithm used the average correlation with the entire velocity field.

Sensors set	"Intuitive" P 3	"Optim" P 3	"Intuitive" P 5	"Optim" P 5
LSE	0.436	0.498	0.532	0.542
QSE	0.58	0.556	0.607	0.626
MTD-LSE	0.697	0.67	0.705	0.695

Table V.3: Determination coefficients R^2 comparison between several estimations using different sets of pressure sensors



a) b)

Fig. V.8: Normalized mean square error of the LSE prediction of the ZDES velocity field using the “optim” P 5 set of sensors. (a) streamwise velocity component, b) vertical velocity component)

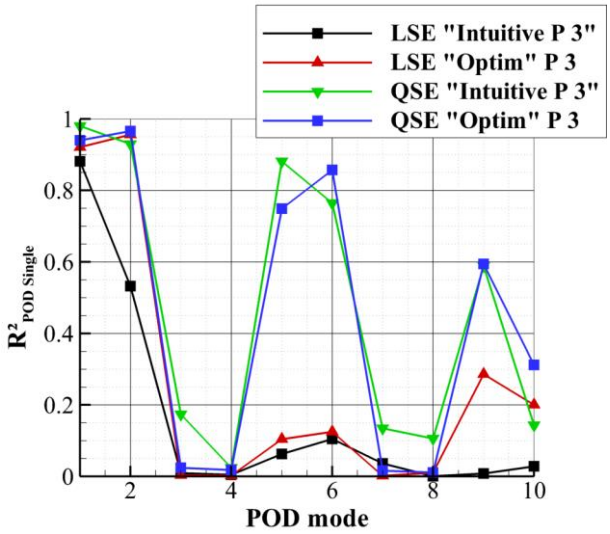


Fig. V.9: Determination coefficient calculated for each POD mode separately on the validation set (OAT15A ZDES)

Determination coefficients of the first 10 POD modes prediction (using LSE and QSE) are plotted in Fig. V.9. The use of the optimized set of sensors improves the prediction of all of the first 10 POD modes, except the seventh. It is consistent with the higher R^2 obtained in this situation (see Table V.3). In the QSE case, the “intuitive” set of sensors performs slightly better for POD modes 3, 7 and 8 than the “optim” P 3 set. That explains the better performances of the “intuitive” P 3 set in QSE.

At last MTD-LSE is investigated. R^2 values (for the best MTD-LSE situations) are reported in Table V.3. These values show that the several sets of sensors perform similarly, with also a very small improvement when two pressure sensors are added. Nevertheless, “intuitive” sets lead to slightly better results than the optimized ones. That is somehow disappointing. But the algorithm does not use delayed information and thus there are no reason why the positions given by the algorithm should lead to the best ones when using MTD-LSE. In addition, it is important to notice that the regularization parameter is chosen using cross-validation with a limited number of possible values for the parameter. Therefore, it may not be perfectly fitted which could participate to the difference observed. For instance, for the “Optim” P 3 set, the regularization parameter (for the best tested MTD-LSE situation) is 1×10^6 and higher and lower tested parameter were 1×10^8 and 1×10^4. Thus, better results might have been obtained with another parameter in this range. In any cases, the differences remain very small less than 3% on the determination coefficient and using the optimized set of sensors do not really penalize the performances.

1.2.2. Backward facing step

In the previous chapter, when studying the length scale estimation by SE, it was observed that the sensor locations had a strong influence on the prediction accuracy of POD modes (see Chapter IV.3.4). In particular, it was shown that POD modes were better estimated when sensors are located near their extrema. This was also observed for the OAT15A airfoil in URANS when using pressure sensors. In Chapter IV.3.4, it was explained that Cohen et al. [99] proposed to choose the sensor at proximity of POD modes extrema. Such method is of course limited when several POD modes have to be considered simultaneously and they proposed to use a cost functional to choose the sensors in such case. Here, we propose to apply the sensor location optimization algorithm previously described.

To apply the algorithm, the sensor space is made of the streamwise velocity information at all mesh points of the PIV snapshots. The data to estimate is one POD coefficient or several. First, the algorithm is used to find positions for one POD mode alone. The first ten sensors chosen by the algorithm for the POD modes 1, 5 and 10 separately are plotted in Fig. V.10 (where the POD mode streamwise velocity component is also shown) and in Fig. V.11 (where the POD mode vertical velocity component is also shown). The locations displayed in these figures match quite well the extremum areas of each POD mode. It confirms that they are the areas of highest correlation between the streamwise velocity data and the POD mode. Also, the chosen locations clearly depend on which POD mode is aimed by the optimization. Therefore, the possibility to apply the algorithm for the estimation of several POD modes simultaneously is particularly helpful.

The algorithm is now used to find 20 streamwise velocity sensors optimized to estimate the first 20 POD modes altogether. Fig. V.12 compares the determination coefficients of the first 50 POD modes using the grids of 33 and 21 sensors (see Chapter IV.3.4), as well as the set of 20 sensors from the optimization algorithm. The LSE using the optimized locations leads to a better reconstruction and prediction of the 20 first POD modes compared 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, see Table IV.3). Moreover, the first two POD modes are predicted with R^2 higher than 80%. It confirms that the location of the sensors plays an important role on the length scales that are predicted by the LSE.

The determination coefficient R^2 of the prediction of the entire flow field, by LSE with the set of 20 sensors (from the algorithm), is about 35%. It 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 IV.2). This shows the effectiveness of the sensor location optimization algorithm, which is able to select a limited number of locations that leads to a better LSE reconstruction and prediction than uniform sensor grids with more sensors.

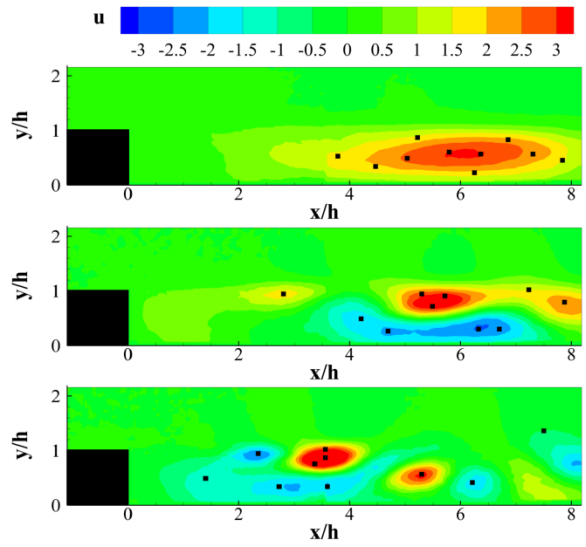


Fig. V.10: Sensor locations obtained by the optimization algorithm for POD mode 1, 5 and 10 separately (from top to bottom). Stream-wise velocity of the modes is displayed.

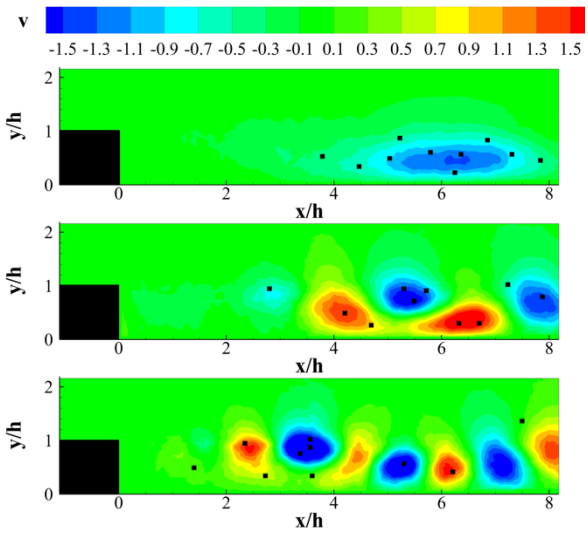


Fig. V.11: Sensor locations obtained by the optimization algorithm for POD mode 1, 5 and 10 separately (from top to bottom). Vertical velocity of the modes is displayed.

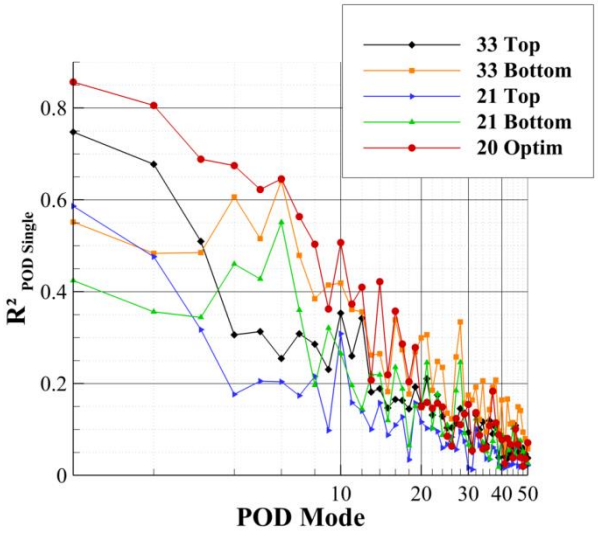


Fig. V.12: Determination coefficient calculated for the validation set as function of the POD mode (BFS using streamwise velocity sensors, low speed PIV database)

At last, another interest of the optimization algorithm is that it can be used to determine the number of sensors necessary to achieve a chosen level of R^2 on the estimation of one POD mode (or of the entire field if the data to estimate is the velocity field and not POD modes). Since the number of sensors will be determined using only training data, it is not guaranteed that this number will indeed be enough for estimation outside the training set. But one can safely expect to obtain a consistent order of magnitude. The number of sensors necessary to reach 50% of $R^2_{POD\ Single}$ for several POD modes are reported in Table V.4. The results confirm that, in the BFS case, more and more sensors are necessary to obtain a good estimation of POD modes when their rank increases. That is consistent with the fact that they possess more and more extrema.

It is now important to notice that the use of the algorithm should be subject to caution when considering POD modes that are not well converged. Indeed, when optimizing the conditional event locations for the prediction of one POD mode only, it was shown that locations close to its extrema

will be chosen. If the POD mode in question is not well converged, then its spatial organization may not be representative of the converged POD mode of same rank. Thus, the chosen locations could be maladjusted for the estimation of data outside the training data set. In the present work, only converged POD modes were considered (the highest POD mode used for the optimization if the 20th POD mode).

POD mode	Number of sensors to reach 50% of R^2
1	2
2	2
5	4
10	7

Table V.4: Number of streamwise velocity sensors required to reach 50% of R^2 using the sensor location optimization algorithm

2. Extension of the sensor location optimization algorithm to the choice of delays

In previous chapters, it was shown that MTD-LSE methods held the best prediction accuracy between the several SE methods tested. The common way to use MTD-LSE is to test several configurations using different number of delays and different time step between them and then to keep the best one. These tests may be time consuming and it would be interesting to dispose of an automatic tool to choose the delays. This is the reason why an extension of the sensor location optimization algorithm is here investigated.

2.1. Description

The extension of the sensor location optimization algorithm to the choice of delays is actually very simple. From a mathematical point of view, one sensor with a delay can be considered as a different sensor and is treated as such in equation (II.10). Therefore, the extension simply consists in testing not only synchronized data at several locations to choose the sensor, but to extend the space of possible sensors to delayed data. The space of possible sensors is thus composed of the synchronized data at every location as well as of the delayed data that one wants to consider. Thus the matrix X_1 becomes:

$$X_1 = \begin{pmatrix} X_1(1)(t) \\ X_1(1)(t-\tau_1) \\ \vdots \\ X_1(1)(t-\tau_{N_d-1}) \\ X_1(2)(t) \\ \vdots \\ X_1(N_S)(t-\tau_{N_d-1}) \end{pmatrix} \in \mathbf{R}^{((N_d \times N_S) \times N)}$$

where N_d is the number of delays considered (synchronous time included).

All the steps of the algorithm remain unchanged.

2.2. Applications

2.2.1. OAT15A airfoil in transonic conditions

- URANS:

In this application, the algorithm is used to choose a set of sensors with delays (from now on, conditional event will be used to refer to a data information at one location with or without delay) optimized for each POD mode separately. In such situation, the locations (as well as the delays) of the chosen conditional events may differ from one POD mode to another. Thus, in the case where one wants to limit itself to use a limited number of locations, the matrix of possibilities X_1 should contain only the data (with all the delays available) at this limited number of locations. For this reason, we limit ourselves to the case where the matrix of possibilities is composed of the pressure data at the 3 locations of the set of sensors "Optim" P 3 with delays, in the past, ranging from 0 to ~1.8 T_{osc} with a time step of 1.7×10^{-5} s (~1.275×10^{-5} T_{osc}).

Table V.5 compares values of R^2 between two cases using the extended optimization algorithm and one case where the set of sensors "Optim P 3" with 80 delays (spaced in time by 6.63×10^{-3} T_{osc}), which corresponds to the best prediction using MTD-LSE-POD among those tested, is used. In this case, denoted "MTD-LSE-POD 80", a total of 240 conditional events is employed and the Tikhonov regularization is used. The two optimized situations correspond to the use of only 3 conditional events ("MTD-LSE-POD Optim 3") and 60 conditional events ("MTD-LSE-POD Optim 60") per POD modes. These conditional events are not necessarily the same for each POD modes. In all cases, only the first four POD modes are estimated. In the optimized situations, the Tikhonov regularization is not applied. In all cases only data from the "Optim P 3" set locations are used. Therefore, the estimation using the extended optimization algorithm does not require any new information than those already used for the "MTD-LSE-POD 80". Reconstruction and prediction results are already quite good when only 3 conditional events chosen by the extended algorithm are used. Prediction in this situation is better than in the case of a straightforward LSE using "Optim P 3" set.

SE methods		LSE-POD ("Optim P 3" set)	MTD-LSE-POD (80 delays every 6.63×10-3 T_{osc})	MTD-LSE-POD (3 optimized conditional events)	MTD-LSE-POD (60 optimized conditional events)
R^2 from original fields	Prediction	0.498	0.929	0.925	0.929
R^2 from POD projected fields	Prediction	0.604	1	0.996	1

Table V.5: Determination coefficient comparison of the MTD-LSE-POD (using four POD modes) in several situations (OAT15A URANS, values from the training set are identical)

$R^2_{\text{POD Single}}$ for the first 10 POD modes are displayed in Fig. V.13 for the training set and in Fig. V.14 for the validation set. The best situation is obtained with "MTD-LSE-POD 80". In this case, all 10 POD modes are estimated with very high fidelity ($R^2_{\text{POD Single}} > 90\%$). Using conditional events chosen by the extended optimization algorithm leads to a deterioration of the estimation of the POD modes when their ranks increase. In the "MTD-LSE-POD Optim 3" case, $R^2_{\text{POD Single}}$ decreases for POD mode of rank higher than 4. If more conditional events are selected by the algorithm, the reconstruction accuracy is improved and the "MTD-LSE-POD Optim 60" case matches the reconstruction accuracy of the "MTD-LSE-POD 80" case. The prediction accuracy of some POD modes is also improved but for other POD modes it is deteriorated. As such, validation set $R^2_{\text{POD Single}}$ for POD modes 9 and 10, in the "MTD-LSE-POD Optim 60" case, are much lower than the training set values. Therefore, using too many conditional events picked by the extended optimization algorithm leads to an overfitted model. In this situation, it is required to use a regularization technique, in a same manner as with the MTD-LSE or MTD-LSE-POD when a lot of delays are used. Applying Tikhonov regularization in such cases could become more complicated than

when only delays at a constant time step are used, because it becomes necessary to find a regularization parameter for each POD mode to estimate. Results using 60 conditional events optimized for each POD mode with and without regularization are shown in Fig. V.15. As expected, prediction accuracy is improved using Tikhonov regularization (regularization parameters are chosen using cross-validation for each POD mode separately). Nevertheless, results are not as good as for the "MTD-LSE-POD 80" but one has to remember that with this test case 240 conditional events are used for each POD mode. Results close to those of the "MTD-LSE-POD 80" case are expected if more conditional events are selected by the extended algorithm and Tikhonov regularization is used.

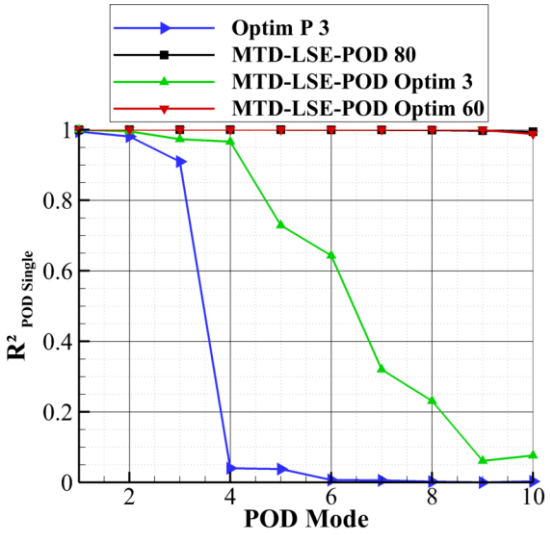


Fig. V.13: Determination coefficient calculated for each POD mode separately for several reconstructions using MTD-LSE-POD (OAT15A URANS)

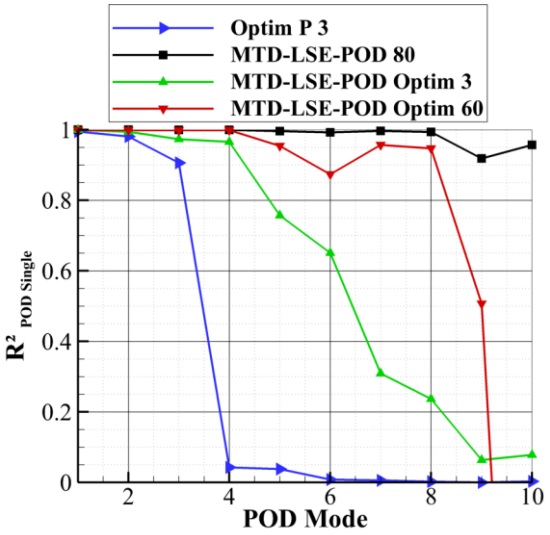


Fig. V.14: Determination coefficient calculated for each POD mode separately for several predictions using MTD-LSE-POD (OAT15A URANS)

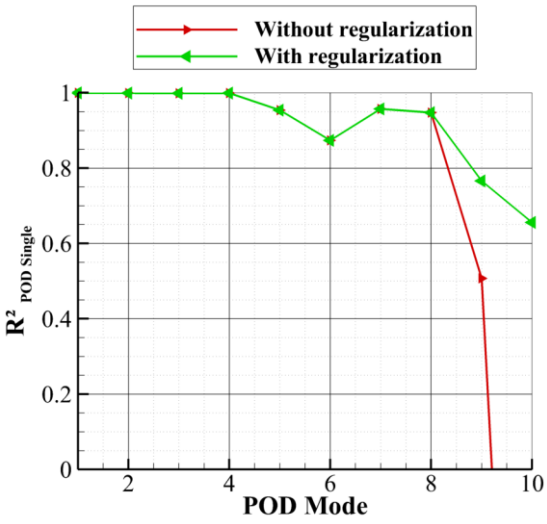


Fig. V.15: Determination coefficient calculated for each POD mode separately using "MTD-LSE-POD Optim 60" with or without Tikhonov regularization (OAT15A URANS)

This application shows that, even without using regularization, the extended algorithm remains able to find a few conditional events that truly improved the reconstruction and prediction compared to single time LSE. In addition, it allows decreasing the size of the problem compared to MTD-LSE and MTD-LSE-POD using delays at constant time step.

- ZDES case:

The extended algorithm is used in a similar fashion as in the URANS case. Conditional events are chosen for each POD mode separately. The space of possible conditional events is limited to the pressure data from the "optim" P 3 set of the ZDES case. Delayed information is added to the space of possible conditional events. Possible delays range from 0 to 1 T_{osc} with a constant step of 1.25×10^{-4} s (~8.62×10^{-3} T_{osc}). Also, only 25 POD modes (about 84% of TKE) are considered.

Determination coefficients of several MTD-LSE-POD configurations are summarized in Table V.6. In one situation, 70 delays are used for all of the 3 sensors of the "Optim P 3" set. This case is denoted "MTD-LSE-POD 70". In addition, Tikhonov regularization is applied in this case. The regularization parameter is chosen using cross-validation. The two other situations use conditional events chosen by the extended sensor location optimization algorithm. In one case, only 3 conditional events are picked for each POD mode (denoted "MTD-LSE-POD Optim 3"). In the second case, 60 conditional events are picked (denoted "MTD-LSE-POD Optim 60"). The best prediction is obtained using the 70 delays for the 3 sensors of the "optim P 3" set. But, "MTD-LSE-POD Optim 3" and "MTD-LSE-POD Optim 60" already lead to a strong improvement of the prediction accuracy compared to the single-time LSE (R^2 is 49.6% on the validation set). R^2 is increased by 10% (on the validation set) by using 3 optimized conditional events and MTD-LSE-POD. In these situations, the number of conditional events used for the estimation of one POD mode is the same as for the estimation using LSE and the "optim P 3" set. That highlights the great interest of using the extended sensor location optimization algorithm to pick delayed information. A great improvement of the reconstruction and prediction accuracy can be obtained without increasing much the number of conditional events and without using regularization.

SE methods		MTD-LSE-POD (70 delays every 8.62×10^{-3} T_{osc})	MTD-LSE-POD (3 optimized conditional events)	MTD-LSE-POD (60 optimized conditional events)
R^2 from original fields	Prediction	0.665	0.597	0.632
R^2 from POD projected fields	Prediction	0.809	0.727	0.77

Table V.6: Determination coefficient comparison of the MTD-LSE-POD (using 25 POD modes and "Optim P 3" set of sensors) in several situations (OAT15A ZDES)

Determination coefficients for the first 10 POD modes estimated using MTD-LSE-POD with the three previous configurations are plotted in Fig. V.16 (training set) and in Fig. V.17 (validation set). These figures show a clear improvement of the reconstruction and prediction of POD modes containing the low frequency content of the flow by using a small number of delays. For instance, with only three conditional events, $R^2_{\text{POD Single}}$ of the POD mode 6 reaches more than 60%, instead of less than 10% when only synchronized information was used. The prediction accuracy in the "MTD-LSE-POD Optim 60" case is not as good as in the "MTD-LSE-POD 70" case. Once again, it is the result of an overfitting of the model that appears when more and more conditional events are chosen by the extended algorithm. The use of Tikhonov regularization, with regularization parameter different for each POD mode and chosen using cross-validation, is investigated and the results are shown in Fig. V.18. The figure demonstrates again that the Tikhonov regularization is indeed able to improve the prediction. R^2 in this case goes from 63.2% to 66.8% which is equivalent to the best MTD-LSE-POD situations (using delays at a constant time step).

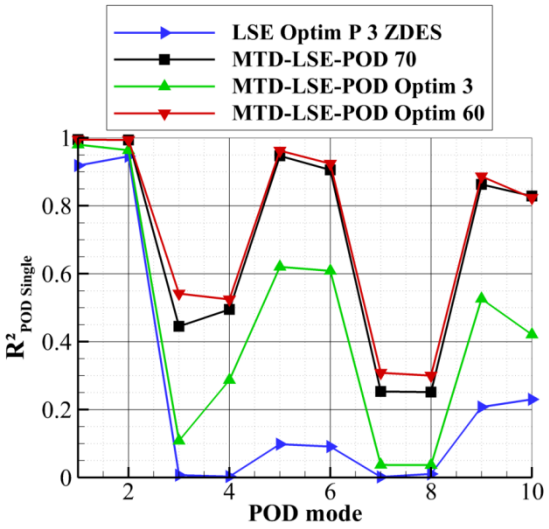


Fig. V.16: Determination coefficient calculated for each POD mode separately for several reconstructions using MTD-LSE-POD (OAT15A ZDES)

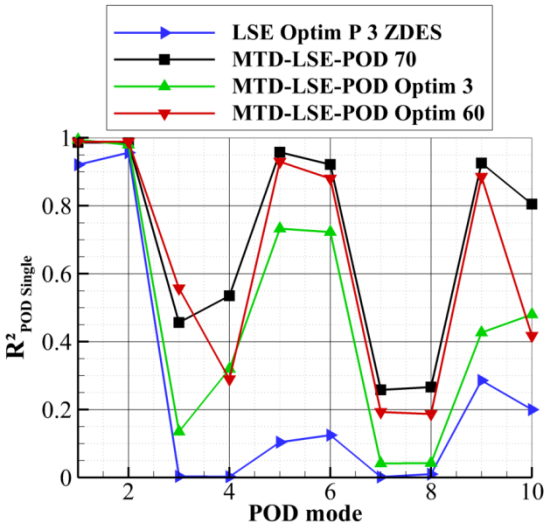


Fig. V.17: Determination coefficient calculated for each POD mode separately for several predictions using MTD-LSE-POD (OAT15A ZDES)

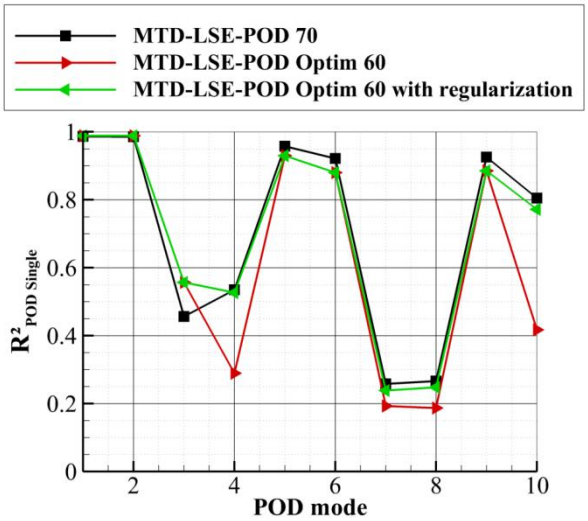


Fig. V.18: Determination coefficient calculated for each POD mode separately using "MTD-LSE-POD Optim 60" with or without Tikhonov regularization. The case where 70 delays are used with the "optim P 3" set and Tikhonov regularization is also plotted. (OAT15A ZDES)

In these examples of applications of the extended sensor location optimization algorithm, this algorithm has in fact been used to optimize the delays to use from an already defined set of sensor locations. In addition, the optimization was performed for POD modes separately. This use of the algorithm shows that it is possible to improve the reconstruction and prediction with a small number of optimized delays. In addition, it is possible to reach similar levels of accuracy than in the best situations using MTD-LSE-POD with constant time step between delays, if Tikhonov regularization is applied. In this case, the number of delays to use is reduced, but it is necessary to find a different regularization parameter for each POD mode.

2.2.2. Backward facing step

In the previous application of the sensor location optimization algorithm in the BFS case, it was shown that the algorithm can be used to determine the number (and the locations) of sensors necessary to achieve a chosen level of accuracy (in terms of R^2) on the estimation (of one or several POD modes or of a velocity field etc.). In the previous chapter, the possibility to improve the accuracy of the reconstruction and prediction using MTD-SE instead of single-time SE has been demonstrated and can be observed in Fig. III.71 or Fig. IV.41 for instance. It is therefore justified to think that the utilization of delays can decrease the number of conditional events necessary to achieve a same level of accuracy. This possibility is now investigated using the high speed PIV database of the BFS test case.

The space of possibilities, for the conditional events, is the entire velocity field (considering only the streamwise velocity component) with 20 delays taken every 5×10^{-4} s (5×10^{-2} $T_{shedding}$) which span a shedding period. The algorithm is use to choose conditional events optimized for separated POD modes. The numbers of conditional events necessary to reach a determination coefficient of 50% for several POD modes are given in Table V.7 ("entire flow field available" columns). "Synchronized only" means that the initial algorithm is used. "Synchronized and delayed" means that the extended algorithm is used. In both case, if the space of possible locations is the entire flow field, the same number of locations is necessary whether or not delayed are considered. Indeed, in this situation, the algorithm chooses mainly sensors synchronized (or with a small delay). Thus, it appears that it is better, in this situation, to use additional synchronized sensors than delayed ones (at an already chosen location). But this does not mean that using MTD-LSE with these optimized locations will not result in better estimation.

In Table V.7, results obtained when the space of possibilities is reduced to a subzone of the initial velocity field are also provided (the subzone is displayed in Fig. V.19). In this situation, considering delayed data becomes interesting. Indeed, less conditional events are required to reach the same level of $R^2_{\text{POD Single}}$ when delayed data can also be chosen. By limiting the region in which sensors can be picked, some "optimal" locations can no longer be reached, and thus more synchronized sensors are required. The delayed data comes in handy as they compensate, for some amount, the synchronized information that is no longer accessible. In overall, there is no downside to use the extended algorithm (even when the space of possible locations is already quite rich) except that the computation time increases with the number of delays considered.

POD mode	Number of conditional events to reach 50% of $R^2_{\text{POD Single}}$			
	Entire flow field available		Subzone available	
	Synchronized only	Synchronized and delayed	Synchronized only	Synchronized and delayed
1	2	2	2	2
2	2	2	2	2
5	5	5	6	5
10	10	10	18	11
20	12	12	24	15

Table V.7: Number of conditional events necessary to reach 50% of $R^2_{\text{POD Single}}$ using only synchronized streamwise velocity data or synchronized and delayed data

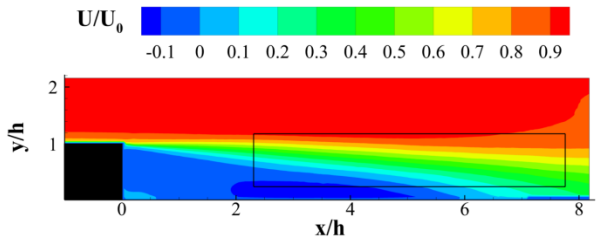


Fig. V.19: Average streamwise velocity field with the reduced subzone for the choice of sensor locations (with or without delays)

3. Modified Stochastic Estimation flow chart

From the previous results, it is possible to draw a flow chart to estimate turbulent flows thanks to SE with POD (see Fig. V.20). 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 to be estimated), the number or the group 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.

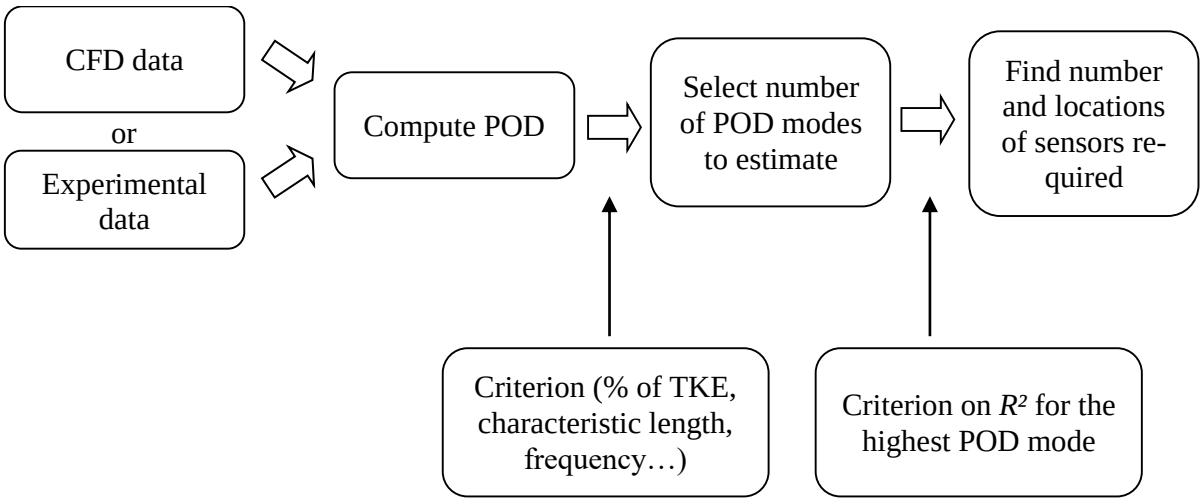


Fig. V.20: Flow chart for the use of SE with POD

4. Chapter summary

The use of the sensor location optimization algorithm proposed by Muradore et al. [98] 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. [99] conclusion. The algorithm has been tested using both the OAT15A database and the BFS database. Its ability to find locations that lead to better reconstruction and prediction accuracy using LSE has been verified and confirmed. When using QSE, some optimized set of sensors did perform with less accuracy than the intuitively chosen ones. And thus, one has to be careful when using other techniques than LSE.

Also, the algorithm represents a convenient tool to determine the sensor locations in order to predict several POD modes simultaneously. It shows the possibility to improve the estimation of some POD modes using the same amount of sensors but with a different placement. Therefore, sensor location plays an important role on the estimation of POD modes and thus on the integral length scales estimated by the LSE (see Chapter IV.3.3 about the relation between length scales and POD modes).

Since it was demonstrated that the use of Multi-Time-Delay LSE (MTD-LSE) improved the estimation accuracy, an extension of the sensor location optimization algorithm has been proposed to choose locations and delays. The goal is to take advantage of the use of delays, without having to try several configurations, by automatically choosing the best delays and locations. The algorithm was first applied to the choice of delays, from a limited number of locations chosen a priori (using the OAT15A databases). The results demonstrated the possibility to improve the prediction accuracy using a small number of delays. In particular, it is possible to obtain the same level of accuracy than in the best situation using constant time step delays. In such case, the use of Tikhonov regularization is however necessary. In the tests performed, no important improvements were observed compared to the use of a large number of delays taken at constant time step. The main advantage is therefore to reduce the size of the problem by decreasing the number of delays used and also to avoid testing several configurations for the delays.

Furthermore, the possibility to decrease the number of conditional events, required to reach a chosen level of accuracy on the estimation, by using synchronized and delayed information was investigated. The results obtained using BFS database show that if the space of possible locations is the entire flow field, then almost only synchronized data are chosen by the algorithm. However, if only a subzone of the flow field can be considered for the locations of the sensors, then considering delayed information decreases the number of conditional events necessary to achieve the same level of accuracy (compared to the case where only synchronized data can be chosen).

Even if the optimization of the sensor locations and delays can improve the SE accuracy, using only wall pressure measurements in the BFS case, the SE prediction quality remains very poor. The improvements obtained with in flow velocity measurements are also limited. Therefore, for such a flow, it appears that other methods than SE ones should be used if one wants to predict the entire velocity field with a high accuracy.

Chapter VI. Study of Kalman Filter for flow prediction

The previous studies of the Stochastic Estimation (SE) have shown that its prediction performances were very limited for some test cases (mainly the Backward Facing Step and the wall mounted cube). The use of multi-time delays methods and of optimized conditional events did not lead to a true breakthrough of the prediction accuracy for these test cases. Therefore, it was decided to investigate an alternative to the SE, the Kalman filter.

In this chapter a short literature overview on Reduced-Order-Modelling in turbulence and data assimilation is first given to introduce the choice of testing the Kalman filter to produce accurate estimate of the velocity field of turbulent flows. Then, the basic equations of the standard Kalman filter but also of the Extended and Ensemble Kalman filter are described. The question of the generation of dynamic and measurements models for the filters is also addressed. At last, the standard, Extended and Ensemble Kalman filter performances are evaluated on three test cases: the URANS simulation of the OAT15A, the ZDES simulation of the OAT15A and the high speed PIV experiment of a wall mounted cube.

1. Short literature overview on Reduced-Order-Modelling and data assimilation

The aim of this literature overview is not to give an exhaustive review on the Reduced-Order-Modelling field or on the data assimilation domain. The goal is here to present some strategies that have been developed in the field of turbulence and that can present an alternative to Stochastic Estimation to provide real time evaluation of flow state.

1.1. Reduced-Order-Model (ROM)

Reduced-Order-Modelling is a very broad subject with applications in many domains. Indeed, Reduced-Order-Modelling has been a necessity for the study of any system which dimensions and complexity made it intractable with the current means of investigation. In a general sense, a ROM is a simplified mathematical model of a complex system that captures most of its physics (or the behavior of interest) while reducing the computational cost of studying the system.

In the field of turbulence, Reduced-Order-Modelling was first considered in order to test or gain physical understandings of turbulent phenomena. One of the first example of ROM, in this field, is the three-mode model of Rayleigh-Bénard convection by Lorenz in 1963 [100]. Using physical approximations, he developed a model formed of three first-order differential equations for the temperature and velocity field dynamics. Later, Aubry et al. developed the first ROM of an open turbulent flow [101]. At the time, another motivation for developing these ROM was the impossibility to numerically simulate the system.

Past and current improvements in computing power now allow the high fidelity simulations of more and more complex systems. But, despite the feasibility of these simulations, the development of ROM remains a critical subject in turbulence. Indeed, ROM can still provide better understandings of the flow physics than the raw simulation results. And they can help to extract useful information from ever larger databases. Also, the development of ROM allows decreasing the computational cost of simulating a complex system, which remains an interesting thing even with the constant increase in computing capacity. Currently, Direct Numerical Simulation is feasible for low Reynolds number and, up to now, the highest Reynolds number of a DNS is $Re_\tau = 5200$ for a wall-bounded turbulent channel flow. DNS is out of reach for an extremely large variety of turbulent flows. Thus, several strategies have been developed in Computational Fluid Mechanics (CFD) to model the turbulence or some scales of the flow (LES, URANS, DES etc.) to realize feasible numerical simulations of turbulent flows. But that is not the focus of this section that concerns the low-order modelization of the system dynamics.

Another reason that makes ROM an important subject in fluid dynamics is the large effort toward the development of active flow control of turbulent flows. Indeed, such control requires the evaluation in real time of several parameters of the system. In a recent publication, Brunton et al. [102] detailed and classified several strategies for the closed-loop control of turbulent systems. In particular, they follow the Wiener classification and thus separate control according to the model used, differentiating ultra-white, white, gray, black box controls and model free controls. Ultra-white box control consists in deriving the control laws analytically from the Navier-Stokes equations. White box control uses CFD (not exclusively DNS but mostly) to supply the information on the system. Gray box control employs ROM of the system dynamics and relates to the underlying physics of the system. Mainly, for turbulent flows, gray box control refers to control using ROM based on modal representation of the system dynamic. Black box control also rests on the utilization of ROM of the system. However, for black box control, ROM are built without much considerations for the underlying physics of the system and mainly from input-output identification. Model free

control does not rely on any model of the system. White box controls are currently far from being feasible for most of real time applications due to their high dimensionality. Model free control is generally limited to open-loop control. Thus, the most common strategies are gray and black box control.

A successfully used black box model in turbulence is the Volterra series ROM. Such models have been used to generate ROM for transonic aerodynamic (see Balajewicz et al. [103]; more details on Volterra series ROM can also be found in Lucia et al. [104]). Recent developments in black box control are also oriented toward Support Vector Machine (see Brunton et al. [102]). But the most popular and investigated category of control for turbulent flows remains the gray box control. Two main methods have been used to form gray box models (see Noack et al. [105]). The first and oldest is the vortex model method. The second is based on Galerkin projection. Vortex models did lead to some successful applications (for instance in a recirculation region, Noack et al. [106]), but they appear to be more complicated to use for control applications compared to Galerkin models. More focus has therefore been given to low-dimensional Galerkin models (Noack et al. [107]).

The basic idea behind the creation of any Galerkin model for turbulent flow is to project the Navier-Stokes equations onto a modal basis, thus obtaining a set of low-order equations in the modal basis. Therefore, an important aspect of Galerkin model is the choice of the modal basis. Several such bases have been developed for turbulent flow. The most common is the Proper-Orthogonal-Decomposition basis (see Chapter II.1.3). The success of the POD in turbulence is directly linked to the popularity of POD Galerkin models (see Chapter I). The first POD ROM of an open turbulent flow was generated by Aubry et al. [101]. Another early example of the use of POD Galerkin projection is the work of Deane et al. [108]. They successfully formed POD ROM for the prediction of limit-cycle behavior of a grooved channel and circular cylinder flows. Since then, POD-Galerkin procedure has also been applied to several turbulent flow configurations such as mixing layer (Ukeiley et al. [109]) and cylinder wake (Noack et al. [110], Mathelin et al. [111]).

While very attractive, the POD-Galerkin method is not without drawbacks. First, as described by Rempfer [112], if the classic Navier-Stokes equations are used, then the knowledge of a pressure term is required on the boundary of the domain investigated. The necessity of a correct evaluation of such term was later characterized more precisely by Noack et al. [105]. In particular, they showed that an appropriate representation of the pressure-term greatly improved the accuracy of POD-Galerkin models. A way to circumvent the need for a correct evaluation of the pressure-term is to use a vorticity formulation of the Navier-Stokes equations. Either way, as it will be recalled in Chapter VI.3.1, the Galerkin projection requires the computation of all inner products and spatial gradients of the flow. Therefore, it can only be performed on numerical data, otherwise strong closure assumptions are necessary. In addition to these limitations, it has been shown that the truncation of the POD basis could lead to inaccurate or instable models. In particular, a stable solution of the Navier-Stokes equations can have an unstable Galerkin model solution (see Rempfer [113] or Noack et al. [114]). Noack et al. [105] also demonstrated that neglecting POD modes of small energy content had a strong influence on the POD-Galerkin model of a two-dimensional mixing layer. A second reason explaining that the POD truncation can lead to diverging models is that low order POD modes mainly resolve the energy production and not so much the energy dissipation in the flow. As a consequence, even if high order POD modes have a low energy contents, they must be taken into account to improve the stability and accuracy of the model. The energy transfers between POD modes of a LES of a backward facing step at $Re_h = 7432$ was investigated by Couplet et al. [115]. They showed that the energy transfer between POD modes was local, followed a forward energy cascade, with some backward cascade. A solution to compensate the missing high order POD modes can be to add some important high order POD modes (see Noack et al. [114]) or to model the influence of the truncated high order POD modes. In particular, a first attempt to counter

the over production of TKE in the POD modes was to add eddy viscosity terms to the model (see Aubry et al. [101], Ukeiley et al. [109], or Couplet et al. [115] for instance).

Another possibility, in order to overcome the necessity of using only numerical data or to model the influence of neglected POD modes, is to use identification methods to compute the POD model parameters. This strategy was proposed by Perret et al. [48]. Couplet et al. [116] also studied the calculation of the POD-ROM parameters through a minimization problem for two simulated flows (wake downstream a square cylinder at Re = 100 and a backward facing step at Re_h = 7432). Le Maître et al. [117] also investigated an equation-free model where the temporal evolution of the POD modes was not obtained through the integration of an ODE. More recently, Cordier et al. [118] investigated a range of identification strategies to estimate the POD galerkin model parameters. In particular, they put in place a data assimilation method, the strong constraints 4D-Var, to improve the POD model parameters.

Before going into the data assimilation subject, it is noteworthy to mention that Galerkin model can be obtained from other modal bases than the POD. Willcox et al. [119] and Rowley et al. [120] have developed an extension of POD model based on a balanced truncation that is called the Balanced POD (BPOD). These models are developed specifically for closed-loop control of turbulent flows. They exploit balanced truncation and approximations of the controllability and observability Grammians of the system. Stability modes of the Navier-Stokes equations can also be used to generate ROM (see for instance Schmid et al. [121]). Linear stability and BPOD were also compared by Tadmor et al. [122]. Another modal basis comes from Dynamic Mode Decomposition (DMD) (Schmid et al. [123]). In DMD, one mode corresponds to one temporal frequency and growth rate. As described by Rowley et al. [124], it is a generalization of spectral analysis to non-linear systems and is a finite-dimensional approximation to Koopman spectral analysis. At last, Kaiser et al. [125] recently developed a new alternative to create ROM called Cluster-based ROM (CROM) and applied it to a two-dimensional incompressible mixing layer as well as to a three-dimensional incompressible turbulent wake of an Ahmed body.

1.2. Data assimilation

Data assimilation refers to a large variety of methods that combine observations of a system with knowledge of its dynamic to provide the best estimate of the state of the system. Data assimilation has been widely developed and applied in meteorological and oceanography fields. A concise introduction to the principle of data assimilation has been written by Talagrand [126]. Data assimilation can be separated into two categories: sequential assimilation and variational assimilation. Sequential assimilation is based on probability considerations and on the propagation in time of the estimate. Sequential assimilation originates from the Bayesian formulation of the assimilation problem and relies on the sequential propagation of the statistics of the vector state. Variational assimilation is based on solving an optimization problem and comes from optimal control theory.

If data assimilation is quite common in meteorology, its use in fluid mechanics is more recent and less spread. Papadakis et al. [127] is among the first to introduce data assimilation in fluid mechanics. In his PhD work, Papadakis applied data assimilation in order to improve the estimation of the parameters of the POD ROM of a cylinder wake flow at Reynolds number 125. In his work, he applied both weak and strong constraints 4D-variational assimilation and showed that the latter lead to the best results. The same kind of approach is followed by Cordier et al. [118] for a mixing layer. Thus, a first use of data assimilation in fluid mechanics is to form more robust ROM by assimilating the ROM parameters. As such, Mons et al. recently applied and compared several data assimilation methods to set the parameters of a dynamic model for the unsteady flow around a cylinder at Re = 100.

Another use is to provide real time estimates of the flow state by combining a ROM (or dynamic model) with observations of the flow. For such applications, variational data assimilation appears to be inadequate and sequential assimilation is used, mainly the Kalman filter (KF). Hœpffner et al. [128] and Chevalier et al. [129] employed KF to estimate the full state of a channel flow in 2005. Later, Sinha et al. [19] developed and compared several estimators for the control of high-speed axisymmetric jets. LSE, QSE, KF, Extended KF (EKF) and linear time-invariant filter (LTIF, in which the Kalman gain is supposed constant) ability to provide the real-time evolution of the POD ROM of the jet is studied. They showed that LTIF performances were identical to the KF one while being as computationally expensive as the LSE. LTIF was found as accurate as QSE and more accurate than LSE. At last, EKF results were not as good as the KF and LTIF ones. Mokhasi et al. [24] investigated the use of unscented KF in order to estimate the simulated flow of a cube mounted in a channel. While the quality of the numerical simulation of the flow is questionable, their work remains particularly interesting in their use of Radial Basis Functions to form a POD ROM of the flow. Ensemble KF (EnKF) was also applied for the estimation of the flow around a NACA0012 airfoil. The filter was combined with a POD ROM obtained for experimental data by Galerkin projection of the Navier-Stokes equations.

From these few applications of KF for turbulent flows, it was decided to investigate the ability of KF to provide more accurate estimation of the flow from limited information than SE methods.

2. Kalman Filter mathematical framework

The basics of the Kalman filter, extended Kalman filter and ensemble Kalman filter are summarized in this section without getting into details for the derivation of the equations. Derivation of the Kalman filter equations can be found in [130], [131] and [132].

As previously stated, Kalman filter is a data assimilation algorithm. It combines observations of the current state of a dynamic system with forecasted state of the system to provide the most accurate estimation of this system. The estimation, given by the Kalman filter, is referred as the analyzed state in the data assimilation community. As described by Talagrand [126], Kalman filter can be deduced from the Best Linear Unbiased Estimate (BLUE, or also called Gauss-Markov theorem) for which the measurement vector is decomposed into two components: a background/forecasted estimate (prior estimator) and an additional set of measurements.

First, some basics of BLUE are recalled. Let us consider some measurements $\boldsymbol{z}(t) \in \mathbb{R}^m$ and a state vector $\boldsymbol{x}(t) \in \mathbb{R}^n$ (n and $m \in \mathbb{N}$). We now want to estimate the state vector using the observations. One way is to consider a linear relationship between the two:

$$\boldsymbol{z}(t) = \Gamma \boldsymbol{x}^{\mathrm{T}}(t) + \boldsymbol{e}(t) \tag{VI.1}$$

Γ is an $m \times n$ matrix that represents a linearization of the relationship between the measurements and the state. $\boldsymbol{e}(t)$ is an error vector due to the linearization and is unknown. However, it is considered to be unbiased ($\mathrm{E}(\boldsymbol{e}) = 0$) and that its covariance matrix is known ($\mathrm{E}(\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}) = \Sigma$ and Σ is known). Then, the Best Linear Unbiased Estimate of $\boldsymbol{x}$ is denoted $\boldsymbol{x}^{\boldsymbol{a}}$ and is defined by (also called the weighted least squares):

$$\boldsymbol{x}^{\boldsymbol{a}}(t) = \mathrm{A}\boldsymbol{z}(t) = (\Gamma^{\mathrm{T}}\Sigma^{-1}\Gamma)^{-1}\Gamma^{\mathrm{T}}\Sigma^{-1}\boldsymbol{z}(t) \tag{VI.2}$$

Now, in the Kalman filter case, the measurement vector is decomposed as (where superscript b stands for background and time is omitted):

$$\boldsymbol{z} = \left(\boldsymbol{x}^{\mathrm{b}}, \boldsymbol{y}\right), \boldsymbol{y} \in \mathbb{R}^p \text{ and } m = n + p$$

In addition, the background state ($\boldsymbol{x}^{\mathrm{b}}$) is assumed to follow (where superscript t stands for the true state):

$$\boldsymbol{x}^{\mathrm{b}} = \boldsymbol{x}^{\mathrm{t}} + \boldsymbol{e}^{\mathrm{b}} \tag{VI.3}$$

with $\boldsymbol{e}^{\mathrm{b}}$ is the error of the background estimate. Also the additional measurement vector is related to the state to estimate by:

$$\boldsymbol{y} = \mathrm{H}\boldsymbol{x}^{\mathrm{t}} + \boldsymbol{\varepsilon} \tag{VI.4}$$

where H is the measurement model matrix and $\boldsymbol{\varepsilon}$ the associated measurement error.

Equations (VI.3) and (VI.4) are equivalent to (VI.1) with $\Gamma = (\mathbf{I}_n, \mathrm{H}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{e} = \left(\boldsymbol{e}^{\mathrm{bT}}, \boldsymbol{\varepsilon}^{\mathrm{T}}\right)^{\mathrm{T}}$. The covariance of the error then becomes:

$$\Sigma = \begin{pmatrix} \mathrm{E}\left(\boldsymbol{e}^{\mathrm{b}}\boldsymbol{e}^{\mathrm{bT}}\right) & \mathrm{E}\left(\boldsymbol{e}^{\mathrm{b}}\boldsymbol{\varepsilon}^{\mathrm{T}}\right) \\ \mathrm{E}\left(\boldsymbol{\varepsilon}\boldsymbol{e}^{\mathrm{bT}}\right) & \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) \end{pmatrix}$$

At last, if background estimate error and the measurement error are uncorrelated, then $\mathrm{E}\left(\boldsymbol{e}^{\mathrm{b}}\boldsymbol{\varepsilon}^{\mathrm{T}}\right) = \mathrm{E}\left(\boldsymbol{\varepsilon}\boldsymbol{e}^{\mathrm{bT}}\right) = 0$. The matrix $\mathrm{E}\left(\boldsymbol{e}^{\mathrm{b}}\boldsymbol{e}^{\mathrm{bT}}\right)$ is denoted P^{b} and is called the background (or forecast) error covariance matrix. $\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})$ is denoted R and is called the measurement model error covariance matrix. Using these notations and hypothesis, the equation (VI.2) can be rewritten as:

$$\boldsymbol{x}^{\mathrm{a}}(t) = \boldsymbol{x}^{\mathrm{b}}(t) + \mathrm{P}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}}\left(\mathrm{H}\mathrm{P}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}} + \mathrm{R}\right)^{-1}\left(\boldsymbol{y} - \mathrm{H}\boldsymbol{x}^{\mathrm{b}}\right) \tag{VI.5}$$

Usually one defines the matrix K called the gain matrix by:

$$\mathrm{K} = \mathrm{P}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}}\left(\mathrm{H}\mathrm{P}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}} + \mathrm{R}\right)^{-1} \tag{VI.6}$$

And the equation becomes:

$$\boldsymbol{x}^{\mathrm{a}}(t) = \boldsymbol{x}^{\mathrm{b}}(t) + \mathrm{K}\left(\boldsymbol{y} - \mathrm{H}\boldsymbol{x}^{\mathrm{b}}\right) \tag{VI.7}$$

The equation (VI.7) is the analysis step of the Kalman filter and describes the correction to apply to the background estimate according to the difference between the measures and the measures that should be obtained if the true state were the background estimate state ($\boldsymbol{y} - \mathrm{H}\boldsymbol{x}^{\mathrm{b}}$). This correction is a linear function of this difference vector that is called the residuals or the innovation vector. In the Kalman filter framework, $\boldsymbol{x}^{\mathrm{a}}$ is called the analyzed state. All the previous equations have been derived for the same instant *t* and thus one can considered that the measurement model matrix H and the measurement model error covariance matrix R are dependent of time. However, in the applications presented in this work, they will be considered constant.

Now let us consider $\boldsymbol{e}^{\mathrm{a}}$ the error between the true state and the analyzed state (called analysis error): $\boldsymbol{e}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{a}} - \boldsymbol{x}^{\mathrm{t}}$, then using equation (VI.4) and (VI.7):

$$\boldsymbol{e}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{b}} + \mathrm{K}\left(\boldsymbol{y} - \mathrm{H}\boldsymbol{x}^{\mathrm{b}}\right) - \boldsymbol{x}^{\mathrm{t}} = \boldsymbol{x}^{\mathrm{b}} + \mathrm{KH}\boldsymbol{x}^{\mathrm{t}} + \mathrm{K}\boldsymbol{\varepsilon} - \mathrm{KH}\boldsymbol{x}^{\mathrm{b}} - \boldsymbol{x}^{\mathrm{t}}$$

$$\boldsymbol{e}^{\mathrm{a}} = \boldsymbol{e}^{\mathrm{b}} - \mathrm{KH}\boldsymbol{e}^{\mathrm{b}} + \mathrm{K}\boldsymbol{\varepsilon} = (\mathrm{I} - \mathrm{KH})\boldsymbol{e}^{\mathrm{b}} + \mathrm{K}\boldsymbol{\varepsilon}$$

Therefore, since the background error and the measurement model error are uncorrelated, the covariance of the analysis error is:

$$\mathrm{P}^{\mathrm{a}} = (\mathrm{I} - \mathrm{KH})\mathrm{P}^{\mathrm{b}}(\mathrm{I} - \mathrm{KH})^{\mathbf{T}} + \mathrm{KRK}^{\mathrm{T}} = (\mathrm{I} - \mathrm{KH})\mathrm{P}^{\mathrm{b}} - (\mathrm{I} - \mathrm{KH})\mathrm{P}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}}\mathrm{K}^{\mathrm{T}} + \mathrm{KRK}^{\mathrm{T}}$$

$$P^a = (I - KH)P^b - (P^bH^T - KHP^bH^T - KR)K^T = (I - KH)P^b - (P^bH^T - K(HP^bH^T - R))K^T$$

Thus, using the expression of K, the analysis error covariance can be simply expressed as:

$$P^a = (I - KH)P^b \tag{VI.8}$$

To complete the description of the Kalman filter, it is now necessary to take into account the way the background estimate is obtained and how the filter is advanced in time.

2.1. Standard Kalman filter

In the case of the standard Kalman filter, the dynamic system is assumed to evolve according to a linear equation. The true state at time t_{k+1} (denoted $\boldsymbol{x}_{k+1}^t$) is linked to the true state at time t_k by (in what follows, subscripts are used to denote the time dependency):

$$\boldsymbol{x}_{k+1}^t = M\boldsymbol{x}_k^t + \boldsymbol{\eta}_k \tag{VI.9}$$

where M is known and called the transition or dynamic model matrix. $\boldsymbol{\eta}_k$ is called the model error and represents processes that are not correctly modelled by M. This error is considered to be an unbiased random vector with a known covariance matrix Q. In addition, the model error is supposed to be uncorrelated with previously introduced errors: $E(\boldsymbol{\varepsilon}_k\boldsymbol{\eta}_l^T) = E(\boldsymbol{e}^a\boldsymbol{\eta}_l^T) = E(\boldsymbol{e}^b\boldsymbol{\eta}_l^T) = 0$.

Now we assume that an analyzed state of the system is available, then the Kalman filter is advanced in time using the transition matrix by:

$$\boldsymbol{x}_{k+1}^b = M\boldsymbol{x}_k^a$$

The error covariance matrix of the predicted state ($\boldsymbol{x}_{k+1}^b$) can then be expressed as function of the analysis error covariance matrix and of the model error covariance matrix. Indeed:

$$P_{k+1}^b = E\left[\left(\boldsymbol{x}_{k+1}^b - \boldsymbol{x}_{k+1}^t\right)\left(\boldsymbol{x}_{k+1}^b - \boldsymbol{x}_{k+1}^t\right)^T\right] = E[(M\boldsymbol{x}_k^a - M\boldsymbol{x}_k^t - \boldsymbol{\eta}_k)(M\boldsymbol{x}_k^a - M\boldsymbol{x}_k^t - \boldsymbol{\eta}_k)^T]$$

$$P_{k+1}^b = MP_k^aM^T + Q \tag{VI.10}$$

Here, matrices M and Q have been considered to be constant, but it is possible to consider time dependent transition matrix and model error without modifying equations (VI.9) and (VI.10). However, in the work presented in this chapter, M and Q will be supposed to be constant over time.

Using the previously explained equations, it is possible to sum up an iteration of the standard Kalman filter algorithm with the following steps:

- Prediction steps:
  - Use the dynamical model to generate a forecast state of the system: $\boldsymbol{x}_{k+1}^b = M\boldsymbol{x}_k^a$
  - Propagate the forecast error covariance matrix: $P_{k+1}^b = MP_k^aM^T + Q$
- Analysis steps:
  - Compute the Kalman gain: $K_{k+1} = P_{k+1}^bH^T\left(HP_{k+1}^bH^T + R\right)^{-1}$
  - Generate the analyzed state: $\boldsymbol{x}_{k+1}^a = \boldsymbol{x}_{k+1}^b + K_{k+1}\left(\boldsymbol{y}_{k+1} - H\boldsymbol{x}_{k+1}^b\right)$
  - Propagate the analysis error covariance matrix: $P_{k+1}^a = (I - K_{k+1}H)P_{k+1}^b$

To initialize the Kalman filter, one has to provide a background state $\boldsymbol{x}_0^{\mathrm{b}}$ and the corresponding background error covariance $\mathrm{P}_0^{\mathrm{b}}$.

2.2. Extended Kalman filter

In the standard Kalman filter, the dynamic model and the measurement model are supposed to be linear. Since most of real systems (in fluid dynamics) have a nonlinear behavior, an extension of the Kalman filter was derived in which the dynamic model ($\mathcal{M}$) and the measurement model ($\mathcal{H}$) are only assumed to be differentiable. To compute the Kalman gain and propagate the error covariance matrices, the Jacobian matrices of *M* and *H* are considered:

$$\mathrm{M}_k = \left.\frac{\partial \mathcal{M}}{\partial x}\right|_{x_{k-1}^{\mathrm{a}}}, \mathrm{H}_k = \left.\frac{\partial \mathcal{H}}{\partial x}\right|_{x_k^{\mathrm{b}}}$$

The algorithm steps are similar to the standard Kalman filter steps:

- Prediction steps
  - Use the dynamical model to generate a forecast state of the system: $\boldsymbol{x}_{k+1}^{\mathrm{b}} = \mathcal{M}(\boldsymbol{x}_k^{\mathrm{a}})$
  - Generate the forecast error covariance matrix: $\mathrm{P}_{k+1}^{\mathrm{b}} = \mathrm{M}_{k+1}\mathrm{P}_k^{\mathrm{a}}\mathrm{M}_{k+1}^{\mathrm{T}} + \mathrm{Q}$
- Analysis steps
  - Compute the Kalman gain: $\mathrm{K}_{k+1} = \mathrm{P}_{k+1}^{\mathrm{b}}\mathrm{H}_{k+1}^{\mathrm{T}}\left(\mathrm{H}_{k+1}\mathrm{P}_{k+1}^{\mathrm{b}}\mathrm{H}_{k+1}^{\mathrm{T}} + \mathrm{R}\right)^{-1}$
  - Generate the analyzed state: $\boldsymbol{x}_{k+1}^{\mathrm{a}} = \boldsymbol{x}_{k+1}^{\mathrm{b}} + \mathrm{K}_{k+1}\left(\boldsymbol{y}_{k+1} - \mathcal{H}(\boldsymbol{x}_{k+1}^{\mathrm{b}})\right)$
  - Generate the analysis error covariance matrix: $\mathrm{P}_{k+1}^{\mathrm{a}} = (\mathrm{I} - \mathrm{K}_{k+1}\mathrm{H}_{k+1})\mathrm{P}_{k+1}^{\mathrm{b}}$

While the extended Kalman filter is a straightforward extension to nonlinear systems, it suffers some limitations. It is common to observe divergence of the extended Kalman filter due to the linearization of the dynamic model and measurement model in the propagation of the error covariance.

To avoid such linearizations, Evensen introduced the ensemble Kalman filter [133].

2.3. Ensemble Kalman filter

We here briefly present the ensemble Kalman filter (EnKF) algorithm. More details about the derivation of the EnKF can be found in [134], [135] and [132]. As stated in the literature review, EnKF is a Monte-Carlo alternative to the extended Kalman filter. The basic idea is to use an ensemble of states, that are propagated by the filter, in order to estimate the error covariances. The EnKF algorithm can be described with the following steps:

- Initialization: from an initial guess of the analyzed state ($\overline{\boldsymbol{x}_0^{\mathrm{a}}}$), $\mathrm{N} \in \mathbb{N}$ random states $\boldsymbol{x}_0^{\mathrm{a},i}$ are generated following a Gaussian law ($\boldsymbol{x}_0^{\mathrm{a},i} \sim \mathcal{N}(\overline{\boldsymbol{x}_0^{\mathrm{a}}}, \mathrm{P}_0)$)
- Prediction steps:
  - Generate N vectors corresponding to the model error: $\boldsymbol{\eta}_k^i \sim \mathcal{N}(0, \mathrm{Q})$
  - Advance in time the N analyzed states: $\boldsymbol{x}_{k+1}^{\mathrm{b},i} = \mathcal{M}\left(\boldsymbol{x}_k^{\mathrm{a},i}\right) + \boldsymbol{\eta}_k^i$

- The forecast state is approximated by the ensemble mean: $\overline{\boldsymbol{x}_{k+1}^{\mathrm{b}}} \approx \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} \boldsymbol{x}_{k+1}^{\mathrm{b},i}$
- The forecast error covariance is approximated by:

$$\mathrm{P}_{k+1}^{\mathrm{b}} \approx \frac{1}{\mathrm{N}-1}\sum_{i=1}^{\mathrm{N}}\left[\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i} - \overline{\boldsymbol{x}_{k+1}^{\mathrm{b}}}\right)\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i} - \overline{\boldsymbol{x}_{k+1}^{\mathrm{b}}}\right)^{\mathrm{T}}\right]$$

- Analysis steps:
  - Generate N vectors corresponding to the measurement error: $\boldsymbol{\varepsilon}_{k+1}^{i} \sim \mathcal{N}(0, \mathrm{R})$
  - Compute the approximate measurement model error covariance:

$$\mathrm{R}_{k+1}^{\mathrm{N}} = \frac{1}{\mathrm{N}-1}\sum_{i=1}^{\mathrm{N}}\left(\boldsymbol{\varepsilon}_{k+1}^{i}\right)\left(\boldsymbol{\varepsilon}_{k+1}^{i}\right)^{\mathrm{T}}$$

  - Kalman gain is computed using the more general expression (relation between equations (VI.6) and (VI.11) is recalled in annex F.1):

$$\mathrm{K}_{k+1} = \mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}}\left(\mathrm{P}_{\boldsymbol{y}\boldsymbol{y}_{k+1}}^{\mathrm{b}}\right)^{-1} \tag{VI.11}$$

    The matrices $\mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}}$ and $\mathrm{P}_{\boldsymbol{y}\boldsymbol{y}_{k+1}}^{\mathrm{b}}$ are then approximated using the ensemble mean:

$$\mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}} \approx \frac{1}{\mathrm{N}-1}\sum_{i=1}^{N}\left[\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i} - \overline{\boldsymbol{x}_{k+1}^{\mathrm{b}}}\right)\left(\mathcal{H}\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}\right) - \overline{\mathcal{H}\left(\boldsymbol{x}_{k+1}^{\mathrm{b}}\right)}\right)^{\mathrm{T}}\right]$$

$$\mathrm{P}_{\boldsymbol{y}\boldsymbol{y}_{k+1}}^{\mathrm{b}} \approx \frac{1}{\mathrm{N}-1}\sum_{i=1}^{N}\left[\left(H\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}\right) - \overline{H\left(\boldsymbol{x}_{k+1}^{\mathrm{b}}\right)}\right)\left(\mathcal{H}\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}\right) - \overline{\mathcal{H}\left(\boldsymbol{x}_{k+1}^{\mathrm{b}}\right)}\right)^{\mathrm{T}}\right] + \mathrm{R}_{k+1}^{\mathrm{N}}$$

  - At last, forecasted states are analyzed using the following formula:

$$\boldsymbol{x}_{k+1}^{\mathrm{a},i} = \boldsymbol{x}_{k+1}^{\mathrm{b},i} + \mathrm{K}_{k+1}\left(\boldsymbol{y}_{k+1} + \boldsymbol{\varepsilon}_{k+1}^{i} - \mathcal{H}\left(\boldsymbol{x}_{k+1}^{\mathrm{b}}\right)\right)$$

    The analyzed state is given by the ensemble mean: $\overline{\boldsymbol{x}_{k+1}^{\mathrm{a}}} = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} \boldsymbol{x}_{k+1}^{\mathrm{a},i}$

3. Dynamic and observation models

As explained in the previous section, Kalman filters require a dynamic model to advance in time the system state and also a measurement model to link the measures to the state. Several methods can be used to provide such models and we describe here those that are investigated in this work. From previous works conducted on ROM for fluid flows, it was decided to consider the system state to be an ensemble of POD coefficients only. The use of other flow decomposition techniques (such as Dynamic Mode Decomposition for instance) to provide a reduced representation of the flow state was not investigated.

Let us denote the reduced system state $\boldsymbol{x}(t) = (a_1(t) \dots a_n(t))^{\mathrm{T}}$ and the observations $\boldsymbol{y}(t) = \left(y_1(t) \dots y_p(t)\right)^{\mathrm{T}}$, we now look for a dynamic model M such as: $\boldsymbol{x}(t+\Delta t) = \mathcal{M}\left(\boldsymbol{x}(t)\right) + \boldsymbol{\eta}(t)$, and a measurement model such as: $\boldsymbol{y}(t) = \mathcal{H}\left(\boldsymbol{x}(t)\right) + \boldsymbol{\varepsilon}(t)$. A conventional way to form the dy-

namic model for the POD coefficients is to use the Galerkin projection of the Navier-Stokes equations onto the POD basis. Concerning the observation models, it is possible to use Gappy POD if the observations are of the same nature than the field used for computing the POD (and measured in the same domain, see Mokhasi et al. [24]). But it is possible to form ad hoc observation models when the observations are not included in the fields used for computing the POD.

3.1. Galerkin model

A Galerkin model is obtained using the standard Galerkin projection of the Navier-Stokes equation onto the selected POD modes forming the reduced system state. To do so, the velocity field in the Navier-Stokes equation (momentum equation) is substituted by its POD decomposition and then the equations are projected onto each POD mode using the dot product between two square integrable velocity fields (see equation (II.13)). Such procedure leads to the following system of Ordinary Differential Equations (ODE) linking the time derivative of each POD coefficient with the other POD coefficients, when the Navier-Stokes equation for a viscous incompressible flow are used:

$$\dot{a}_i(t) = \nu \sum_{j=0}^{n} l_{ij}^{\nu}\, a_j(t) + \sum_{j,k=0}^{n} q_{ijk}^{c}\, a_j(t) a_k(t) + f_i^{p} \tag{VI.12}$$

where:

$i \in [\![0, n]\!]$, the first POD mode corresponds to the average field and its POD coefficient is constant equal to 1.

$l_{ij}^{\nu} = \left(\boldsymbol{\Phi}^{(i)}, \Delta\boldsymbol{\Phi}^{(j)}\right)_{\Omega}$ is the dissipation (or also called viscous) term

$q_{ijk}^{c} = -\left(\boldsymbol{\Phi}^{(i)}, \left(\boldsymbol{\Phi}^{(j)} \cdot \boldsymbol{\nabla}\right)\boldsymbol{\Phi}^{(k)}\right)_{\Omega}$ is the convection term

$f_i^{p} = -\frac{1}{\rho}\left(\boldsymbol{\Phi}^{(i)}, \boldsymbol{\nabla} P\right)_{\Omega}$ is the pressure term

The derivation of this equation is recalled in annex G.

In the case where one has only access to the velocity field, the pressure term cannot be obtained directly and has to be modelled. Noack et al. [105] showed that this term has to be correctly taken into account and that for open flows this term could be represented by a constant linear fit with the POD coefficients. As such the pressure term can be approximated by:

$$f_i^{p} \approx \sum_{j=1}^{n} l_{ij}^{p} a_j(t)$$

Another problem rising from POD-ROM using Galerkin projection is that only a limited number of POD modes are considered and resolved. Since the most energetic POD modes are usually conserved, they better resolve the production than the dissipation of TKE which can lead to a diverging model. It is therefore important to correctly model the effect of the truncation. As described in the literature overview, one popular way of solving the problem is to add an eddy-viscosity, but such models require a very careful calibration. Following these remarks, it was decided to circumvent those challenges by directly forming a dynamic model (of the form of equation (VI.12)) without directly calculating the POD Galerkin projection.

More generally, following equation (VI.12), it appears that the dynamic model can be written as, $\forall i \in [\![1, n]\!]$:

$$\dot{a}_i(t) = c_i + \sum_{j=1}^{n} l_{ij}\, a_j(t) + \sum_{j,k=1}^{n} q_{ijk}\, a_j(t) a_k(t) \tag{VI.13}$$

Or in a discrete form, using a first order scheme:

$$a_i(t + \Delta t) = a_i(t) + \dot{a}_i(t)\Delta t$$

$$a_i(t + \Delta t) = C_i + \sum_{j=1}^{n} L_{ij}\, a_j(t) + \sum_{j,k=1}^{n} Q_{ijk}\, a_j(t) a_k(t) \tag{VI.14}$$

with:

$C_i = c_i \Delta t$,

$L_{ij} = \begin{cases} l_{ij}\Delta t \text{ if } i \neq j \\ l_{ii}\Delta t + 1 \text{ if } i = j \end{cases}$,

$Q_{ijk} = q_{ijk}\Delta t$.

Three methods are considered to form the dynamic model: the LSE, the QSE and a Radial Basis Function (RBF) model. In addition, four-dimensional data assimilation (4D-Var) is also considered to improve the model when LSE or QSE is used. LSE was already used to form a dynamic model of a flow by Tu et al. [65]. RBF has also been used by Mokhasi et al. [24]. For all three methods, it is necessary to have time-resolved information of the velocity field or at least sampled at a sufficiently high frequency so that for all the POD coefficients *i* considered in the reduced order model, $a_i(t + \Delta t)$ and $a_i(t)$ are not completely uncorrelated. For the observation model, no attempt was made to physically draw such a model and LSE or QSE are also used. The derivation of models using LSE or QSE is identical to the method used by Perret et al. [136]. In his work however, they also considered cubic terms in the ODE forming the POD-ROM.

3.2. LSE models

3.2.1. Dynamic model

In the case where LSE is used to form the dynamic model, then only the linear term in equation (VI.14) is considered. Quadratic and constant terms are thus set to zero.

To form a dynamic model using LSE, the data to estimate are composed of POD coefficients (considered for the ROM) at time $t + \Delta t$ and the sensors are composed of the same coefficients at time *t*. Therefore, if a succession of *N* temporal samples is available for the POD coefficients, then the LSE is trained using the following matrices for the data to estimate (D) and the sensors (S):

$$\mathrm{D} = \begin{pmatrix} a_1(t_2) & \cdots & a_1(t_N) \\ \vdots & \ddots & \vdots \\ a_n(t_2) & \cdots & a_n(t_N) \end{pmatrix} \in \mathbb{R}^{n\times(N-1)}, \mathrm{S} = \begin{pmatrix} a_1(t_1) & \cdots & a_1(t_{N-1}) \\ \vdots & \ddots & \vdots \\ a_n(t_1) & \cdots & a_n(t_{N-1}) \end{pmatrix} \in \mathbb{R}^{n\times(N-1)}$$

The coefficients of the linear dynamic model are obtained using equation (II.8) where:

$$\mathrm{A} = \begin{pmatrix} L_{11} & \cdots & L_{n1} \\ \vdots & \ddots & \vdots \\ L_{1n} & \cdots & L_{nn} \end{pmatrix}, \mathrm{C} = \mathrm{S}\mathrm{S}^{\mathrm{T}}, \mathrm{B} = \mathrm{S}\mathrm{D}^{\mathrm{T}}$$

At last, in order to be used with the Kalman filter, it is necessary to dispose of the model error covariance matrix. This matrix is estimated using a validation set of data on which the error is calculated between the true state and its LSE prediction (using the true state at the precedent time step as sensor).

3.2.2. Observation model

To generate the observation model by LSE, it is simply assumed that the observations are linearly related to the reduced order state, $\forall i \in [\![1, p]\!]$:

$$\boldsymbol{y}_i(t) = \sum_{j=1}^{n} L_{ij}^{\mathrm{obs}} a_i(t)$$

The model coefficients (L_{ij}^{obs}) are obtained by LSE considering the data matrix D and the sensor matrix S are:

$$\mathrm{D} = \begin{pmatrix} \boldsymbol{y}_1(t_1) & \cdots & \boldsymbol{y}_1(t_N) \\ \vdots & \ddots & \vdots \\ \boldsymbol{y}_p(t_1) & \cdots & \boldsymbol{y}_p(t_N) \end{pmatrix} \in \mathbb{R}^{p \times N}, \mathrm{S} = \begin{pmatrix} a_1(t_1) & \cdots & a_1(t_N) \\ \vdots & \ddots & \vdots \\ a_n(t_1) & \cdots & a_n(t_N) \end{pmatrix} \in \mathbb{R}^{n \times N}$$

Similar to the dynamic model derived by LSE, the observation model error covariance matrix can be computed through the estimation of the error between the true observation and the observation predicted by LSE from the true state on a validation set.

3.3. QSE models

3.3.1. Dynamic model

QSE dynamic model is obtained by considering the linear and quadratic terms of equation (VI.14), the constant coefficients are set to zero. The coefficients of the model are obtained by training the QSE with the same data and sensors as for the LSE dynamic model. The dynamic model error covariance matrix is evaluated in a similar manner on a validation set.

3.3.2. Observation model

QSE observation model is formed by considering the following equation relating the observations to the state:

$$\forall i \in [\![1, p]\!], \boldsymbol{y}_i(t) = \sum_{j=1}^{n} L_{ij}^{\mathrm{obs}} a_i(t) + \sum_{j=1}^{n} \sum_{k=1}^{n} Q_{ijk}^{\mathrm{obs}} a_i(t) a_j(t)$$

The coefficients L_{ij}^{obs} and Q_{ijk}^{obs} are obtained by training the QSE with the observations as data to estimate and the reduced order state as sensors. The observation model error covariance matrix is evaluated on a validation set as explained for LSE models.

3.4. Four dimensional variational data assimilation

The variational data assimilation used is based on the work of Papadakis [127], who introduced data assimilation in fluid mechanics, as summarized by Cordier et al. [118]. Only the strong constraint four-dimensional variational data assimilation (4D-Var) is considered and used to assimilate coefficients of the dynamic model. The main formulae are described here, precise derivation of those equations can be found in [127] (chapter 3).

The dynamic system considered is the one described in equation (VI.13) which can be written as:

$$\frac{d\boldsymbol{x}(t)}{dt} + \mathcal{M}(\boldsymbol{x}(t), \boldsymbol{C}, \boldsymbol{L}, \boldsymbol{Q}) = 0 \tag{VI.15}$$

$\mathcal{M}$ is the nonlinear operator relative to the dynamic model, $\boldsymbol{C}$ is the vector of constant coefficients, $\boldsymbol{L}$ is the matrix of linear coefficients and $\boldsymbol{Q}$ is the tensor of quadratic coefficients. In addition, a constant initial conditions for the system state is assumed: $\boldsymbol{x}(t_0) = \boldsymbol{x}^0$.

The goal of the strong constraint assimilation is to find the solution $\boldsymbol{x}(t)$ (of equation (VI.15) with the initial condition: $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$) that best fits some observations $\boldsymbol{y}(t)$ over a finite time interval $[\![0, T]\!]$. Here, the only way to modify the solution $\boldsymbol{x}(t)$ is to change the parameters c_i, l_{ij}, q_{ijk} of the dynamic model. Thus the problem is to find the dynamic model parameters that minimize the following cost functional:

$$\begin{aligned}\mathcal{J}(\boldsymbol{C}, \boldsymbol{L}, \boldsymbol{Q}) = \frac{1}{2}\int_0^T \|\boldsymbol{y}(t) - H\boldsymbol{x}(t)\|^2 \\ + \frac{\alpha}{2}(\|\boldsymbol{C} - \boldsymbol{C}^0\|^2 + \|\boldsymbol{L} - \boldsymbol{L}^0\|_F^2 + \|\boldsymbol{Q} - \boldsymbol{Q}^0\|_F^2)\end{aligned} \tag{VI.16}$$

$\|.\|_F$ denotes the Frobenius norm. α is a regularization parameter (not considered in the work of Papadakis, but used by Cordier et al. [137]). In our case, the observations to assimilate will directly be the POD coefficients obtained by computing the POD on the training data set. Therefore, the cost functional becomes:

$$\mathcal{J}(\boldsymbol{C}, \boldsymbol{L}, \boldsymbol{Q}) = \frac{1}{2}\int_0^T \left\|\boldsymbol{x}^{\text{obs}}(t) - \boldsymbol{x}(t)\right\|^2 + \frac{\alpha}{2}(\|\boldsymbol{C} - \boldsymbol{C}^0\|^2 + \|\boldsymbol{L} - \boldsymbol{L}^0\|_F^2 + \|\boldsymbol{Q} - \boldsymbol{Q}^0\|_F^2)$$

Superscript 0 stands for the initial guess of the dynamic model parameters. These parameters can be directly obtained by POD Galerkin projection, or by linear regression (such as LSE or QSE).

Since the determination of the gradient of the cost functional is not practically feasible (using finite difference) for large systems, the minimization problem is solved using the adjoint formulation. It is possible to show that the forward integration in time of the dynamic system (VI.15) followed by a backward integration of an adjoint dynamical model allows to determine the cost functional gradient. Demonstration is given by Papadakis in [127]. Using a Lagrangian functional, the adjoint dynamical model can be expressed as:

$$-\frac{\partial \boldsymbol{\theta}}{\partial t}(t) + \left(\frac{\partial \mathcal{M}}{\partial \boldsymbol{x}}\right)^* \boldsymbol{\theta}(t) = \boldsymbol{x}^{\text{obs}}(t) - \boldsymbol{x}(t) \tag{VI.17}$$

$\theta(t) \in \mathbb{R}^n$ is a Lagrange multiplier. * stands for the adjoint operator and is here equivalent to conjugate transpose. $\frac{\partial \mathcal{M}}{\partial \boldsymbol{x}}$ is the linear tangent of the dynamic model $\mathcal{M}$:

$$\left(\frac{\partial \mathcal{M}}{\partial \boldsymbol{x}}\right)_{ij} = l_{ij} + \sum_{k=1}^{n} \left(q_{ikj} + q_{ijk}\right) a_k(t)$$

At last the optimality conditions are:

$$\begin{aligned} \frac{\partial \mathcal{J}}{\partial c_i} &= \alpha(c_i - c_i^0) + \int_0^T \theta_i(t)\, dt = 0 \\ \frac{\partial \mathcal{J}}{\partial l_{ij}} &= \alpha\left(l_{ij} - l_{ij}^0\right) + \int_0^T \theta_i(t) a_j(t)\, dt = 0 \\ \frac{\partial \mathcal{J}}{\partial q_{ijk}} &= \alpha\left(q_{ijk} - q_{ijk}^0\right) + \int_0^T \theta_i(t) a_j(t) a_k(t)\, dt = 0 \end{aligned} \tag{VI.18}$$

Equations (VI.15), (VI.17) and (VI.18) form the optimality system. To find the optimal parameters of the dynamic model, this system is solved iteratively. First, the dynamic system is integrated forward in time (equation (VI.15)), then the adjoint equation is integrated backward in time (equation (VI.17)) and at last the dynamic model parameters are updated (equations (VI.18)). After that the three previous steps are repeated using the updated coefficients as initial guess. In this work, since a first order forward scheme is considered to form the discrete dynamic model used in the Kalman filter, the time integration of equations (VI.15), (VI.17) and in (VI.18) are performed using the same scheme.

3.5. Radial Basis Function (RBF) models

3.5.1. Dynamic model

The RBF model is formed using a network of RBF to directly link the state at time $t + \Delta t$ with the state at time *t*:

$$\begin{aligned} &\forall i \in [\![1, n]\!], a_i(t + \Delta t) = \mathrm{RBF}_{\mathrm{network}}^i(\boldsymbol{x}(t)) \\ &\boldsymbol{x}(t) = \begin{pmatrix} \mathrm{RBF}_{\mathrm{network}}^1(\boldsymbol{x}(t)) \\ \vdots \\ \mathrm{RBF}_{\mathrm{network}}^n(\boldsymbol{x}(t)) \end{pmatrix} \end{aligned} \tag{VI.19}$$

An RBF network is a sum of RBF that generally takes the form:

$$\mathrm{RBF}_{\mathrm{network}}^i\left(\boldsymbol{x}(t)\right) = \sum_j \lambda_j^i \Phi^i\left(\left\|\boldsymbol{c}_j - \boldsymbol{x}(t)\right\|\right)$$

where:

$\|.\|$is the Euclidian norm in $\mathbb{R}^n$, $\boldsymbol{c}_j \in \mathbb{R}^n$ are called the centers, $\Phi^i(r)$ are the radial basis function and $\lambda_j^i \in \mathbb{R}$ are coefficients that need to be computed.

Usually the radial basis functions are chosen to be positive definite functions in order to lead to a well-defined interpolation problem. Common functions used are the Gaussian function ($\Phi(r) = \mathrm{e}^{-\epsilon^{i^2}\|r\|^2}$), the multi-quadratic function ($\Phi(r) = \sqrt{1 + \epsilon^{i^2}\|r\|^2}$) or the inverse multi-quadratic

function ($\Phi(r) = \frac{1}{\sqrt{1+\epsilon^{i^2}\|r\|^2}}$). Some other functions are sometimes considered such as polyharmonic spline or thin plate spline. The parameter ϵ^i is called the shape parameter or scaling parameter and is a very important parameter to tune the RBF network. It is possible to use a single parameter for all the RBF in the sum, or to use a different parameter for each RBF. In this work, only the first option is considered. A way to determine this or these coefficients is to use Leave-One-Out Cross-Validation (LOOCV) (once the centers have been chosen). This method is used in this work. More precisions on LOOCV can be found in [138] (in particular the algorithm 1 is used).

The first step to form the RBF model is to choose the centers $\boldsymbol{c}_j$. One simple choice, when one is using a training set of data to form the model, is to use all the data from the training set as centers. In our case, we dispose of *N*-1 values for which the relation $\boldsymbol{x}(t+\Delta t) = M(\boldsymbol{x}(t))$ is known, therefore the centers are chosen to be $\boldsymbol{x}(t_k), \forall k \in [\![1, N-1]\!]$. This procedure was used by Mokhasi et al. [24]. Another possibility is to perform unsupervised learning such as k-means to found some centers inside the training data set [72].

Using all the training data as centers and after the shape parameters have been fixed, it is possible to easily obtain the λ_j^i coefficients by interpolation. Indeed, for $i \in [\![1, n]\!]$ and for $j \in [\![1, N-1]\!]$ we search λ_j^i such as:

$$\forall k \in [\![1, N-1]\!], \sum_{j=1}^{N-1} \lambda_j^i \Phi^i\left(\left\|\boldsymbol{x}(t_j) - \boldsymbol{x}(t_k)\right\|\right) = a_i(t_{k+1})$$

Which can be rearranged as:

$$\begin{pmatrix} a_i(t_2) \\ \vdots \\ a_i(t_N) \end{pmatrix} = \begin{pmatrix} \Phi^i(\|\boldsymbol{x}(t_1) - \boldsymbol{x}(t_1)\|) & \cdots & \Phi^i(\|\boldsymbol{x}(t_{N-1}) - \boldsymbol{x}(t_1)\|) \\ \vdots & \ddots & \vdots \\ \Phi^i(\|\boldsymbol{x}(t_1) - \boldsymbol{x}(t_{N-1})\|) & \cdots & \Phi^i(\|\boldsymbol{x}(t_{N-1}) - \boldsymbol{x}(t_{N-1})\|) \end{pmatrix} \begin{pmatrix} \lambda_1^i \\ \vdots \\ \lambda_{N-1}^i \end{pmatrix}$$

$$\mathrm{A}^i = \Phi^i \Lambda^i$$

This system has to be solved for each POD coefficients forming the state. Fortunately, with this choice of centers, and since the RBF are positive definite functions, the matrix Φ^i is symmetric positive definite and thus can be inverse.

The procedure used to form the RBF model can be summarized as follow. Gaussian functions are chosen as RBF. All the data from the training set are used as centers for the RBF. Then several shape parameters are tested consecutively for each member of the system state (each POD coefficient) in order to choose the best ones by LOOCV. At each test, all the coefficients λ_j^i have to be computed.

3.5.2. Observation model

The observation model using RBF is obtained in the same manner than the dynamic one. Simply the equation (VI.19) becomes:

$$\boldsymbol{y}(t) = \begin{pmatrix} \mathrm{RBF}^1_{\mathrm{network}}(\boldsymbol{x}(t)) \\ \vdots \\ \mathrm{RBF}^p_{\mathrm{network}}(\boldsymbol{x}(t)) \end{pmatrix}$$

4. Applications

In this section, Kalman filter and its variants are applied to some test cases used to evaluate the performances of SE. First, the Kalman filter is applied to the OAT15A URANS and ZDES simulations. Then, it is used on the data from the high speed PIV campaign of the wall mounted cube.

4.1. OAT15A URANS test case

The estimation using SE-POD of the OAT15A URANS flow field has been conducted in Chapter IV.2. Ten POD modes were considered and this amount of POD modes contained about 98% of TKE. When SE is used to estimate a POD coefficient, the number of POD coefficients considered does not impact the quality of the estimation of each POD mode individually. However, using Kalman filter, the size of the state considered (and thus the number of POD coefficients retained) directly impacts the dynamic model and the Kalman filter results. Therefore, four state sizes are compared. 4 (93% TKE), 10 (97.6% TKE), 50 (99.93% TKE) and 100 (99.99% TKE) POD coefficients are used.

As explained in Chapter II.4.2.1, the training set is formed of an oscillation period and the validation set of the following oscillation period. Only the set of 3 pressure sensors which locations have been optimized ("Optim P 3") is used (see Chapter V.2.2.1.). Filters are initialized using the LSE of the state at the first time step.

4.1.1. Standard Kalman filter

Using the standard Kalman filter (KF), only linear dynamic model and linear observation model are used. Both are constructed using LSE to identify the model coefficients. The accuracy of both models is evaluated using the determination coefficient calculated for each estimated data separately following equation (II.27). Determination coefficients are shown in Fig. VI.1 and Fig. VI.2 for the observation and dynamic models respectively. For the dynamic model, POD modes higher than 50 are not shown but their R^2 remains higher than 95%. In overall, all models perform very well. In addition, using more and more POD coefficients improves the accuracy of both observation and dynamic models.

Forward in time integration of the dynamic model from the same initialization as the one of the Kalman filter is performed. This study allows appreciating the impact of assimilating the observations (in the KF) by comparing these results with those of the KF. It also gives some information on the dynamic model behavior and mainly about its stability property. The results show that the four dynamic models are stable. Surprisingly, adding more POD modes does not improve the quality of the prediction obtained by the straightforward integration of the model. The determination coefficients of the first 50 POD modes, obtained on the validation set, are plotted in Fig. VI.3. Only the first four POD modes are predicted with a positive R^2, all the other POD modes are badly predicted. In addition, it is clear that for the first four POD modes, R^2 decreases when more POD modes are considered for the dynamic model. This tendency is opposite to the one obtained when looking at the prediction of the state at a time step from the true state at the previous step. This indicates that the dynamic model becomes more sensible to noise (or more generally error) when more POD modes are used. Indeed, its prediction accuracy improves if the true state is used. But it decreases when a perturbed state is used. The initial error in the initial state used for the integration has a greater impact when the dynamic model is formed using POD modes. Even if the initial condition of the integration is the true state, the same evolution as in Fig. VI.3 is observed. Therefore, it is not the consequence of an increasing error in the initial condition when the state size increases.

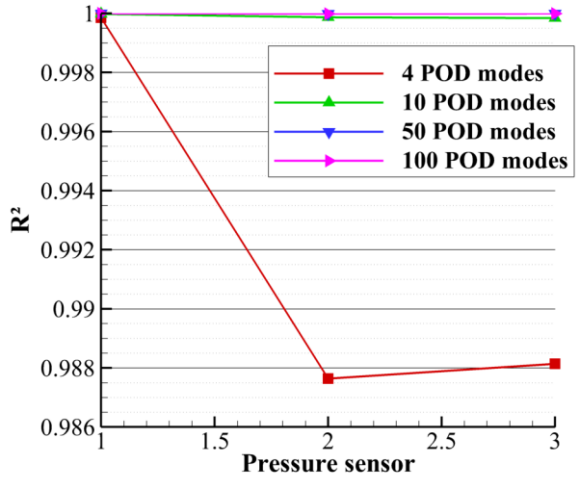


Fig. VI.1: Determination coefficients of the estimation, using LSE, of the 3 pressure sensors (from "Optim P 3" set) using 4, 10, 50 and 100 POD modes as measures (observation model)

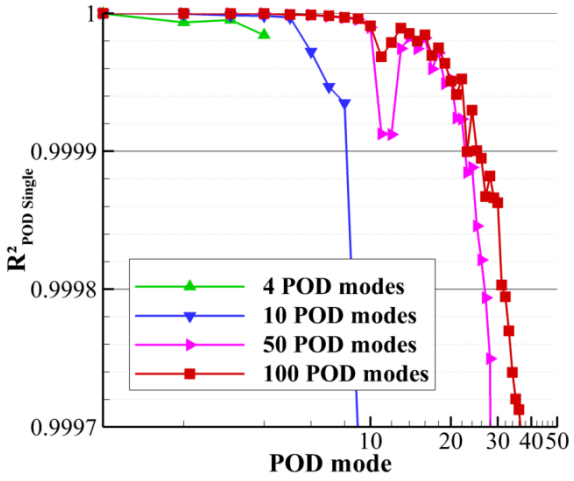


Fig. VI.2: Determination coefficients of POD coefficients estimated by LSE using 4, 10, 50 and 100 POD coefficients at the previous time step as measures (dynamic model)

Fig. VI.4 shows the determination coefficient of the first 10 POD modes estimated using the standard KF in the four situations. Comparing the results with those in Fig. VI.3, it is clear that the assimilation of the observations as a positive impact on the prediction accuracy. $R^2_{\text{POD Single}}$ are higher for the Kalman filter than for the forecast using the dynamic model only.

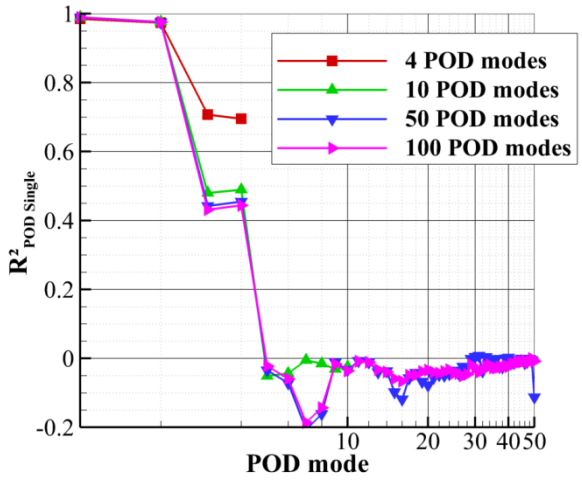


Fig. VI.3: Determination coefficient of the POD mode prediction by integration of the dynamic model (validation set)

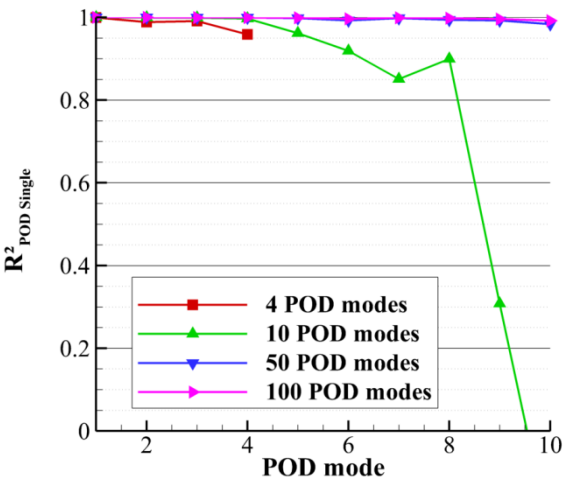


Fig. VI.4: Determination coefficient of POD coefficients estimated using standard Kalman filter and "Optim P 3" set of pressure sensor (validation set)

Fig. VI.4 can also be compared with Fig. V.5. From this comparison, it is clear that the standard KF performs better that the LSE. However, the results are changed when different numbers of POD modes are considered to form the system state. In this situation, the use of 100 POD modes leads to the best estimation using the standard KF. Results obtained for the estimation of the complete flow field are compared in Table VI.1. It is clear that even using only 4 POD modes, the

standard KF leads to a better estimation than the LSE. Using 10 POD modes, the results are as good as the QSE and using 100 POD modes they are of the same quality (and slightly better) than the best MTD-LSE result. These results show that it is possible to improve the estimation using KF instead of SE. However, it is important to remind that KF is clearly more computationally expensive than SE.

Since the standard KF is stable in the 4 tested situations, a way to decrease its computational cost is to consider a constant Kalman gain K. Indeed, the Kalman gain converged when the KF is stable. Using a constant Kalman gain obtained after a first run of the Kalman filter leads to a very slight worsening in the present case. Using 100 POD modes, R^2 (with the entire flow field as reference) goes from 99.62% to 99.56%.

Type of estimation		R^2 (%)
Standard Kalman filter	4 POD modes	92.5
	10 POD modes	96.59
	50 POD modes	99.03
	100 POD modes	99.62
LSE		88.74
QSE		96.18
MTD-LSE		99.41

Table VI.1: Determination coefficients comparison between standard Kalman filter and SE ("Optim P 3" set of pressure sensors is used)

4.1.2. Extended Kalman Filter

The use of a quadratic dynamic model with the extended Kalman filter (EKF) is now investigated. The observation model remains however linear. EKF is known to suffer from divergence due to a bad propagation of the error covariance matrices. In the present case, careful tuning of the dynamic model (obtained by QSE) through Tikhonov regularization was required in order to obtain a converging EKF. Choosing the regularization parameter, using Cross Validation (CV) directly when forming the model by QSE proved to be ineffective and CV has to be applied to the EKF results instead. Thus, for several Tikhonov parameters it was necessary to form the dynamic model and then to run the EKF to be able to choose the best parameter. For all cases, the resulting dynamic model prediction abilities (by time integration) are greatly decreased compared to the LSE linear dynamic model. The results rapidly converged toward zero for all coefficients forming the system state (see Fig. VI.9 and Fig. VI.10, curve with Tikhonov parameter α equal to 5).

Determination coefficients for the prediction of the entire flow field, using EKF, are summarized in Table VI.2. These values can be compared with the ones in Table VI.1 and it is clear that the use of EKF with a quadratic dynamic model obtained by QSE does not lead to a better estimation of the flow. Even more interesting, the prediction is deteriorated when the number of POD coefficients used to form the state increases. Two factors participate to these observations. First, due to an incorrect propagation of the model error covariance using linear tangent, higher than optimal (according to CV) regularization parameters are necessary to form QSE dynamic models which leads to a converging EKF. Then, using these too high regularization parameters, the resulting QSE dynamic models are actually less accurate than the LSE ones. This is illustrated in Fig. VI.7 where $R^2_{\text{POD Single}}$ of the prediction of the first 10 POD coefficients using LSE and QSE dynamic models (and the same coefficients at the previous time step as measurements) are shown. The QSE dynamic model used with the EKF is the one with $\alpha = 5$. It is clear that this dynamic model is less accurate than the LSE one.

Type of estimation		R^2 (%)
EKF	4 POD modes	90.5
	10 POD modes	89.7
	50 POD modes	89.3
	100 POD modes	88.7

Table VI.2: Determination coefficients for the prediction of the entire velocity field using EKF with a quadratic model from QSE and the sensor set "Optim P 3"

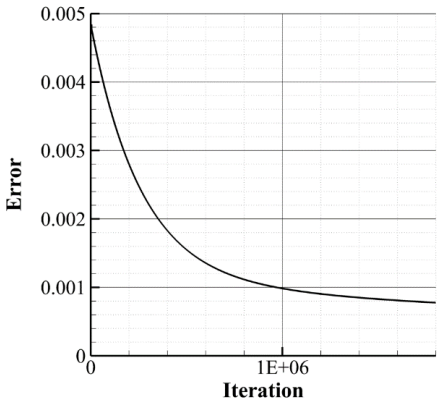


Fig. VI.5: Error between the assimilated data and the original one as function of the algorithm iteration (assimilation of 4 POD coefficients)

At last, a quadratic dynamic model obtained using 4D-Var data assimilation is quickly investigated only when 4 POD coefficients are used. The model is obtained by initializing the assimilation algorithm with the coefficients of the linear dynamic model obtained by LSE. And the assimilation is conducted over the first period forming the training data set. The algorithm was iterated 1 800 000 times. The evolution of the error defined by $Error = \int_0^T \left\| \boldsymbol{x}^{\text{obs}}(t) - \boldsymbol{x}(t) \right\|^2$ is plotted in Fig. VI.5. R^2 obtained for the prediction of the flow field is about 92.5% which is approximately the same value obtained with the standard KF.

In the several tested situations, no improvements were obtained using EKF instead of standard Kalman filter. Therefore, the use of the Ensemble Kalman filter (EnKF) is investigated.

4.1.3. Ensemble Kalman Filter

The use of the Ensemble Kalman filter requires choosing a number of statistical samples (ensemble members) that will be randomly drawn to represent the error statistics of the system state. In our case, the largest state considered contained 100 POD coefficients, and thus it was decided to use 1 000 ensemble members. Using EnKF offer a large freedom concerning the dynamic and observation models used. First, to validate the EnKF implementation, linear dynamic and observation model obtained by LSE are utilized and the results compared with those of the standard Kalman filter.

4.1.3.1. Linear dynamic and observation model

Fig. VI.6 compares the determination coefficients of the prediction of the first 10 POD modes using KF and EnKF with linear dynamic and observation models from LSE (system state is composed of the first 10 POD modes). This figure clearly shows that both filters perform the same with negligible difference as it was expected.

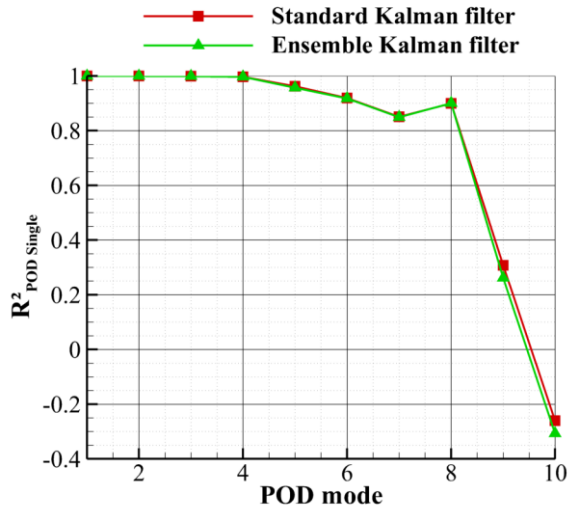


Fig. VI.6: Comparison of determination coefficients of the first 10 POD modes prediction using KF and EnKF with linear dynamic and observation model

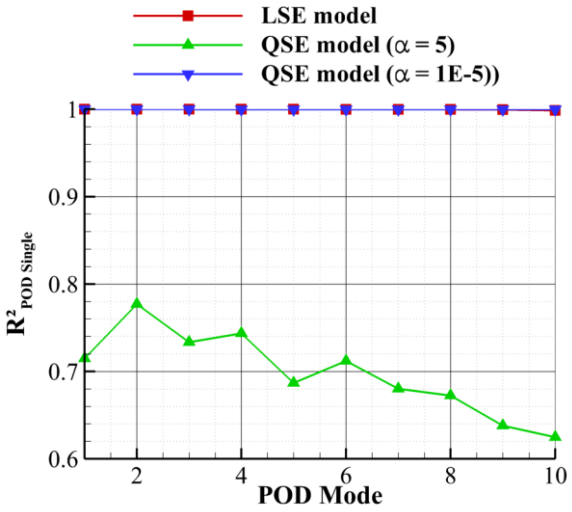


Fig. VI.7: Determination coefficients of POD coefficients estimated by LSE and QSE using 10 POD coefficients at the previous time step as measures

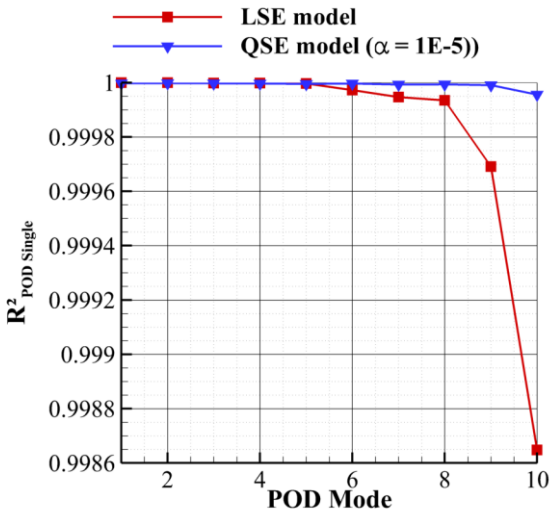


Fig. VI.8: Determination coefficients of POD coefficients estimated by LSE and QSE using 10 POD coefficients at the previous time step as measures

4.1.3.2. Quadratic dynamic and linear observation model

The EnKF is used with a quadratic dynamic model formed using QSE. The observation model is linear and formed by LSE. Compared to the EKF, the EnKF seems to be indeed more stable. As such, using a state composed with the first 4 POD coefficients, no Tikhonov regularization is necessary when forming the dynamic model to obtain convergence of the filter. However, as the number of POD coefficients used as system state increases, regularization becomes necessary. Compared to the EKF, Tikhonov parameters are lower to achieve the best prediction. Also the values obtained by applying CV when creating the QSE model or by applying CV on the EnKF results are similar. As such, with 10 POD coefficients, α is equal to 1×10^{-5} to obtain convergence and the best prediction using EnKF, instead of 5 with the EKF. The resulting QSE dynamic model is then more accurate than the LSE one (see Fig. VI.8). Looking at the forecasting ability of these models (by

time integration), they appear to be unstable and they rapidly diverge. Results are illustrated in Fig. VI.9 and Fig. VI.10. In these figures, the prediction of the first and tenth POD coefficients for the two oscillation periods are plotted. Linear and quadratic models are formed using 10 POD coefficients for the system state. The results displayed indicate that the linear dynamic model is the one producing the best forecast over time. The quadratic model used for the EKF corresponds to $\alpha = 5$ and the one used for the EnKF to $\alpha = 1\times10^{-5}$. The first one rapidly converges to zero after 3 iterations; the second one diverges after 5 iterations.

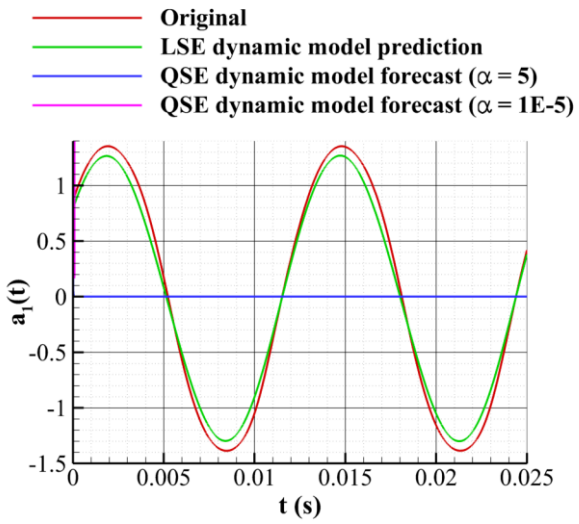


Fig. VI.9: Forecast of the first POD coefficient using linear and quadratic dynamic models

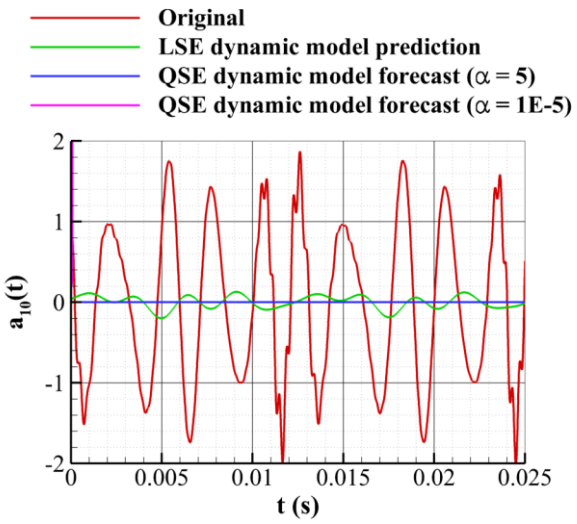


Fig. VI.10: Forecast of the tenth POD coefficient using linear and quadratic dynamic models

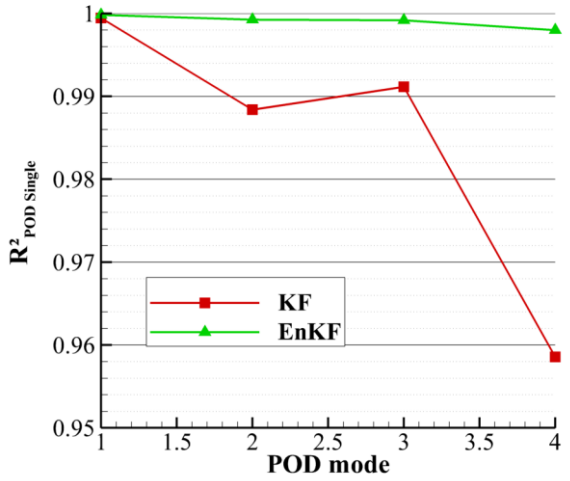


Fig. VI.11: Comparison of determination coefficients of the first 4 POD modes prediction using KF and EnKF with quadratic dynamic model and linear observation model (4 POD coefficients system state)

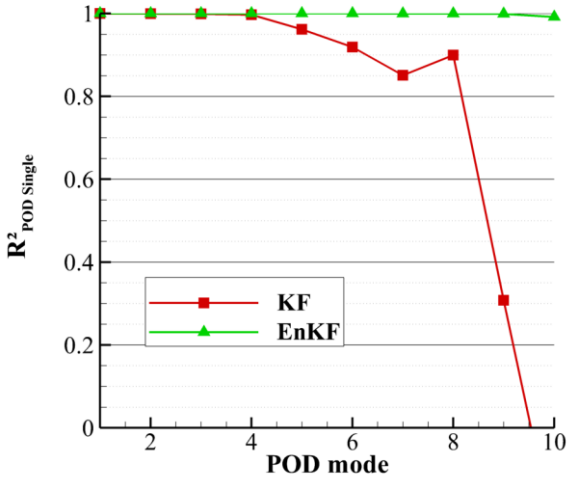


Fig. VI.12: Comparison of determination coefficients of the first 10 POD modes prediction using KF and EnKF with quadratic dynamic model and linear observation model (10 POD coefficients system state)

Results of the EnKF are plotted in Fig. VI.11, Fig. VI.12, Fig. VI.13, Fig. VI.14 and compared with the standard KF ones. Except when the system state is made of the first 100 POD coefficients, EnKF demonstrates better prediction performances than the standard KF for all the estimated POD

coefficients. On the opposite, when 100 POD coefficients are used, the EnKF leads to a deterioration of the prediction quality compared to the standard KF. That is due to the overfitting of the model that becomes too important. As such, even with regularization, the QSE dynamic model is less accurate than the LSE one. In addition, the best QSE dynamic model (obtained by CV to choose the regularization parameter) leads to a diverging EnKF and the best EnKF results are obtained for a higher regularization parameter (therefore for an even less accurate quadratic dynamic model) which explained the deterioration compared to the use of a linear dynamic model.

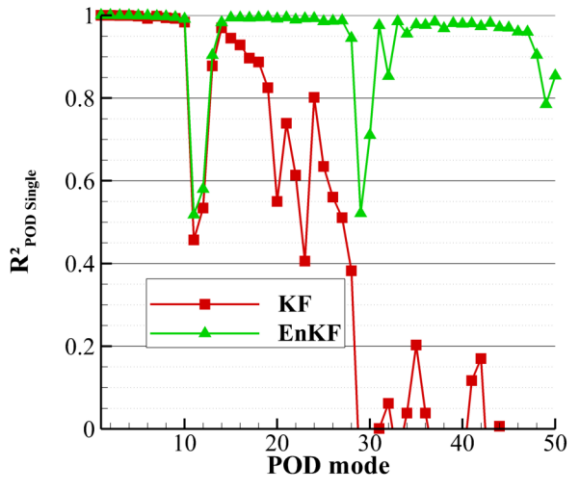


Fig. VI.13: Comparison of determination coefficients of the first 50 POD modes prediction using KF and EnKF with quadratic dynamic model and linear observation model (50 POD coefficients system state)

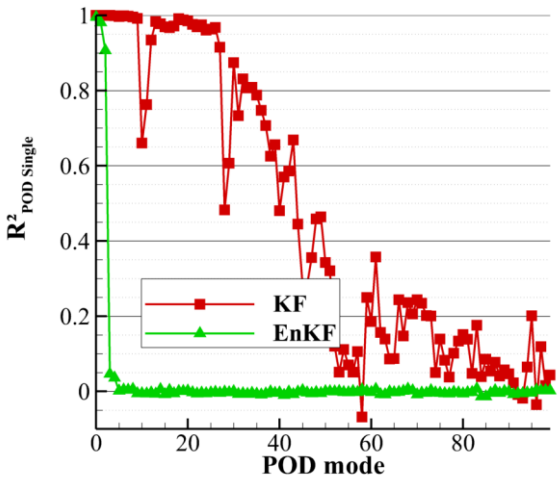


Fig. VI.14: Comparison of determination coefficients of the first 100 POD modes prediction using KF and EnKF with quadratic dynamic model and linear observation model (100 POD coefficients system state)

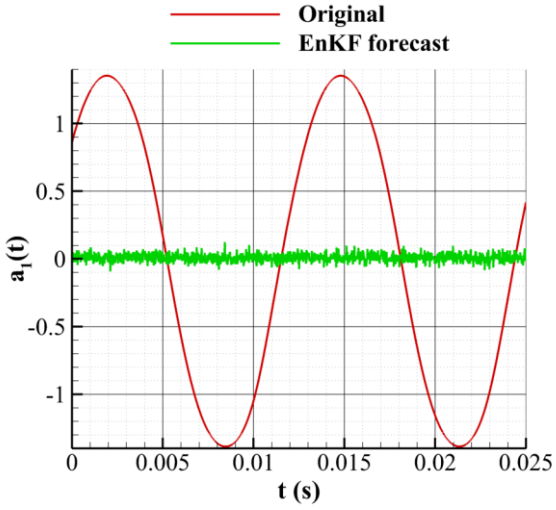


Fig. VI.15: Forecasted first POD coefficient in the EnKF using a state of 100 POD coefficients

Actually, the EnKF results in this situation are very close to the one obtained with LSE-POD. This observation suggests that the dynamic model is, in some sense, rapidly disregarded by the EnKF (since it is too inaccurate compared to the observation model) and that the EnKF results come almost entirely from the assimilation of the pressure sensors information. That is confirmed by

looking at the forecasted state generated at the prediction steps of the EnKF algorithm (see Chapter VI.2.3). The forecast of the first POD coefficient is plotted in Fig. VI.15 as an example. It is clear that the forecasted state is close to zero at all time. In such conditions, the analysis step is almost equivalent to the LSE using the set of pressure sensors. (The relation between the analysis step and LSE is explored in annex F.3). As expected, the tuning of the dynamic model is obviously a critical parameter with a strong impact on the KF performances.

The previous observations are confirmed by the R^2 values obtained for the velocity field prediction that are reported in Table VI.3. R^2 is slightly higher using EnKF instead of KF in the 3 first cases and deteriorates when 100 POD coefficients is used.

Type of estimation		R^2 (%)
EnKF	4 POD modes	92.9
	10 POD modes	97.58
	50 POD modes	99.61
	100 POD modes	88.74

Table VI.3: Determination coefficients for the prediction of the entire velocity field using EnKF with quadratic dynamic models from QSE and the sensor set “Optim P 3”

In the present case, no QSE dynamic model was able to lead to a good EnKF when the system state is made of 100 POD coefficients. So, the possibility to obtain a better quadratic dynamic model using data assimilation was investigated. The assimilation algorithm was run for 200 000 iterations. The determination coefficient calculated with the velocity flow field as reference is about 99.44%. This is higher than R^2 obtained with the QSE dynamic model but remains lower than the standard KF. In the present case, it might be possible to improve this result using more iterations for the assimilation algorithm, but it is likely that the results would not be significantly better than in the standard KF. Indeed, standard KF results are already almost perfect. The main result here is to show that the data assimilation can serve to form quadratic dynamic models that are more robust and accurate than those formed by QSE with Tikhonov regularization.

Following these results, the impact of using a quadratic observation model is studied.

4.1.3.3. Linear dynamic model and quadratic observation model

In the previous section, the impact of using a quadratic dynamic model instead of a linear one has been observed. In this section, we look at the possibility to use a quadratic observation model formed by QSE with a linear dynamic model from LSE in the EnKF. For the four test cases, the best results are obtained for observation models from QSE without regularization.

Results are presented in Table VI.4. As in the case where quadratic dynamic model is used instead of linear dynamic model, improvement of prediction accuracy is obtained in the situation using the less POD coefficients in the system state. A slight increase of R^2 is obtained in the 4 and 10 POD modes cases, on the contrary R^2 decreases for 50 and 100 POD modes cases. This deterioration comes from the overfitting of the quadratic dynamic model and in particular from the conditioning of the matrix made of the system states that are considered as sensors to form the observation QSE model. As the number of POD coefficients considered increases, the condition number drastically increases too. If the use of Tikhonov regularization limits the effect of such bad conditioning, it however leads eventually to an observation model that is less accurate than the linear one obtained by LSE (in a similar fashion to what occurs for the QSE dynamic model in the 100 POD coefficients state case). Thus, the EnKF results are less accurate when a QSE observation model is used with system state containing 50 and 100 POD coefficients.

A possibility to avoid this problem is to use an "inverse" observation model. Such method was used by Mokhasi et al. [24]. What we call "inverse" observation model corresponds to the use of an estimation of the state as measurement in the analysis step of the KF algorithm. Thus instead of using: $\boldsymbol{y} = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{\varepsilon}$, $\boldsymbol{x} = \mathcal{H}'(\boldsymbol{y}) + \boldsymbol{\varepsilon}$ is used. The modification of the KF formula are described in annex F.2. The EnKF results using "inverse" QSE quadratic observation model are summarized in Table VI.5. In addition, the results obtained with "inverse" LSE linear observation model are also displayed. These results can be respectively compared with those in Table VI.4 and Table VI.1 respectively. Using "inverse" linear observation models gives similar results than using direct linear observation models but there is a small discrepancy for 50 and 100 POD coefficients case. None of these models suffer from severe overfitting. However, there is no guarantee that both "direct" and "inverse" models should lead to the same EnKF results since they are different models. Concerning the quadratic observation model, using an "inverse" model clearly improves the results in the 50 and 100 POD coefficients cases compared to the use of a "direct" model. This is due to the fact that contrary to the direct model, the "inverse" model does not suffer from overfitting. Indeed, the inverse model employs less conditional events than the direct model (especially when many POD coefficients are used for the system state). Therefore, the condition number of the conditional events matrix is less likely to be very high. However, the "inverse" model results remain lower than standard KF ones for 50 and 100 POD coefficients cases.

Type of estimation		R^2 (%)
EnKF	4 POD modes	92.87
	10 POD modes	97.41
	50 POD modes	89.34
	100 POD modes	88.89

Table VI.4: Determination coefficients for the prediction of the entire velocity field using EnKF with linear dynamic models from LSE, quadratic observation model from QSE and the sensor set "Optim P 3"

Type of estimation		"inverse" LSE observation model	"inverse" QSE observation model
EnKF	4 POD modes	92.36	92.92
	10 POD modes	96.08	97.25
	50 POD modes	98.33	98.97
	100 POD modes	98.64	99.21

Table VI.5: Determination coefficients for the prediction of the entire velocity field using EnKF with linear dynamic models from LSE and "inverse" observation models

4.1.3.4. Quadratic dynamic and observation models

Using quadratic dynamic and observation models leads to a complicated tuning of the EnKF. Indeed, one has to choose regularization parameters for both models (when necessary). Once more CV on the EnKF results was used to set the regularization parameters of the two QSE models. The results are given in Table VI.6, as previously "direct" and "inverse" observation models are tested.

For the 4 and 10 POD modes cases, both "direct" and "inverse" observation model lead to the same results of the EnKF prediction and none of them suffer from overfitting. For the two other cases, using an "inverse" observation model improves the EnKF prediction and is therefore a better solution than Tikhonov regularization applied to the direct model. For the 4, 10 and 50 POD modes cases, the EnKF results combining quadratic dynamic and observation models are the most accurate ones among the KF tested situations. They are also better than the LSE results. The best prediction

is obtained when the state is made of 50 POD modes and the EnKF prediction in this situation is also more accurate than the QSE and MTD-LSE prediction.

In the case where 100 POD modes formed the system state, the EnKF results using QSE dynamic and observation model are deteriorated compared to the case where a LSE dynamic model is used. As previously explained, it is due to the fact that the QSE dynamic model is less accurate than the LSE one. In addition, when a "direct" QSE observation model is used, the results are very poor. This illustrates how a bad tuning of the EnKF can turn out to have disastrous results while the goal was to improve the accuracy by increasing the complexity of both dynamic and observation models.

Type of estimation		"direct" QSE observation model	"inverse" QSE observation model
EnKF	4 POD modes	92.95	92.95
	10 POD modes	97.59	97.59
	50 POD modes	99.53	99.73
	100 POD modes	38.67	98.64

Table VI.6: Determination coefficients for the prediction of the entire velocity field using EnKF with QSE dynamic and observation models

4.1.3.5. Radial Basis Function dynamic model

The use of RBF dynamic model is investigated only on the 4 and 100 POD coefficients cases for brevity. In both cases, the centers of the RBF functions are the 749 first system state samples of the training set (the training set is made of 750 and the last one is not considered as center since the model estimate the system state from the previous one in time). As explained in Chapter VI.3.5 shape parameters are obtained by LOOCV.

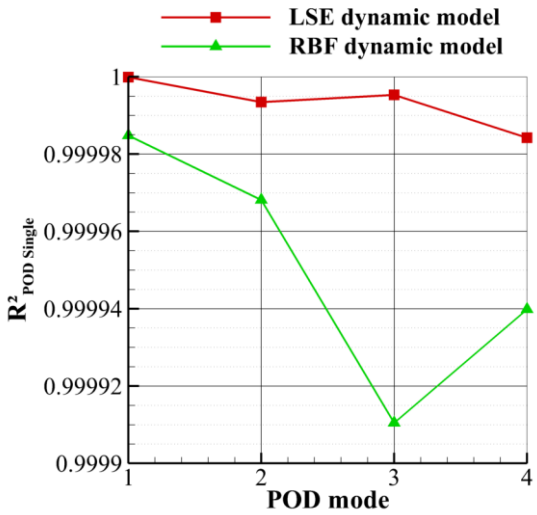


Fig. VI.16: Determination coefficients of POD coefficients estimated by LSE and RBF using 4 POD coefficients at the previous time step as measures (validation set)

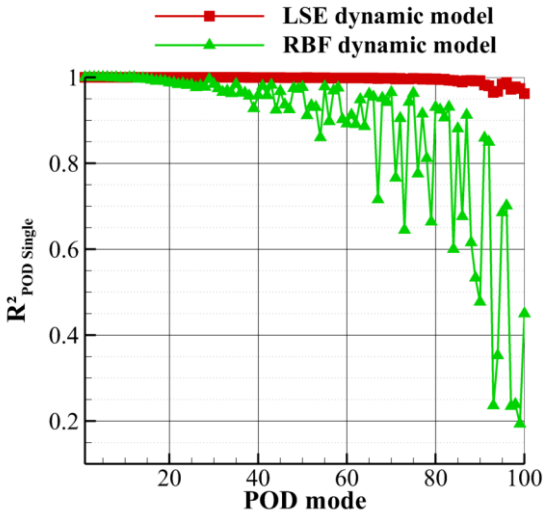


Fig. VI.17: Determination coefficients of POD coefficients estimated by LSE and RBF using 100 POD coefficients at the previous time step as measures (validation set)

For both 4 and 100 POD coefficients case, the RBF dynamic model obtained is less accurate than the linear one created by LSE. Determination coefficients of the prediction of the 4 and 100

POD coefficients using the POD coefficients at the previous time step as measurements are displayed in Fig. VI.16 and Fig. VI.17 (R^2 is calculated for the validation set). LSE and RBF dynamic model are compared. The figures show that for all POD coefficients, the LSE dynamic model performs better than the RBF one. Therefore, Kalman filter is expected to have better prediction results when LSE dynamic model is used than RBF.

The results of the EnKF with the RBF dynamic model (and LSE observation model) are summarized in Table VI.7 and compared with the results obtained from standard KF with LSE dynamic model. R^2 is clearly deteriorated by the use of the RBF dynamic model. When 100 POD coefficients form the system state, R^2 even becomes negative with the RBF model. In the present conditions, LSE or QSE dynamic models perform better than RBF ones.

Type of estimation	EnKF with RBF dynamic model	Standard KF with LSE dynamic model
4 POD modes	18.94	92.5
100 POD modes	-4.09	99.62

Table VI.7: Determination coefficients (in %) for the prediction of the entire flow field using EnKF with RBF dynamic model and using standard KF

4.2. OAT15A ZDES case

It was shown in Chapter IV.2 that the number of POD modes, necessary to retain the same amount of TKE, was higher in the OAT15A ZDES case than in the URANS case. Therefore, the numbers of POD coefficients considered to form the system state are chosen higher than for the URANS case. 10, 52, 205 and 1000 POD coefficients states are studied. They correspond respectively to 75.8%, 90.1%, 97.6% and 99.9% of TKE. Since it was previously shown that EKF does not hold any significant interest compared with the EnKF, only EnKF is used when dynamic or observation models are not linear. Two sets of pressure sensors are used. They are the sets “Optim P 3” and “Optim P 5” of the ZDES case (see Chapter V.2.2.1).

4.2.1. Standard Kalman filter

The quality of the linear dynamic models obtained by LSE in the four situations is quickly compared. $R^2_{\text{POD Single}}$ is displayed for the first 205 POD modes for the LSE prediction using the POD coefficients at the previous time step as measurements in Fig. VI.18. Also, up to POD mode 1000, $R^2_{\text{POD Single}}$ higher than 50% are obtained. The accuracy of those dynamic models is satisfying for low rank POD modes but is clearly lower for POD modes of ranks higher than 200. In addition, contrary to what was observed in the OAT15A URANS case, increasing the number of POD coefficients considered in the system state (and so in the model) does not constantly improve the model accuracy. The results show that the best model, at least the more accurate for POD modes of ranks 1 to 150, is the model obtained with 205 POD coefficients to form the state. In these conditions, taking into account more and more POD coefficients in the Kalman filter, using LSE dynamic model, may deteriorate the prediction ability of the filter.

R^2 of the “Optim P 5” pressure sensor prediction using 10, 52, 205 and 1000 POD coefficients and LSE is plotted in Fig. VI.19 (observation model). Except for the first sensor, for which the prediction accuracy of LSE observation models from 10 and 52 POD coefficients is less than 85%, all LSE observation models perform quite well ($R^2 > 90\%$). The observation models obtained using 205 and 1000 POD coefficients have almost the same accuracy, the second one performing slightly better.

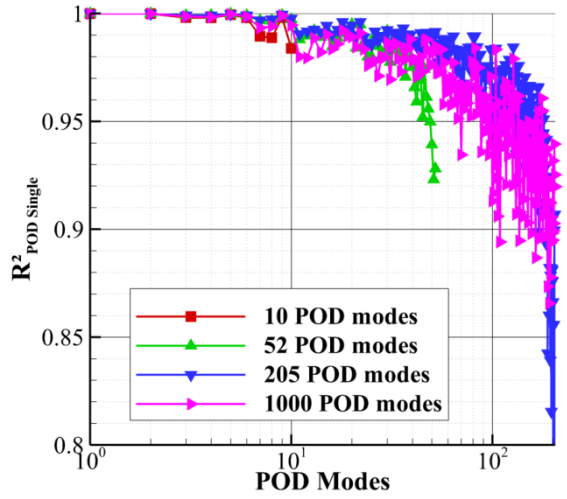


Fig. VI.18: Determination coefficients of POD coefficients estimated by LSE using 10, 52, 205 and 1000 POD coefficients at the previous time step as measurements (validation set, dynamic model)

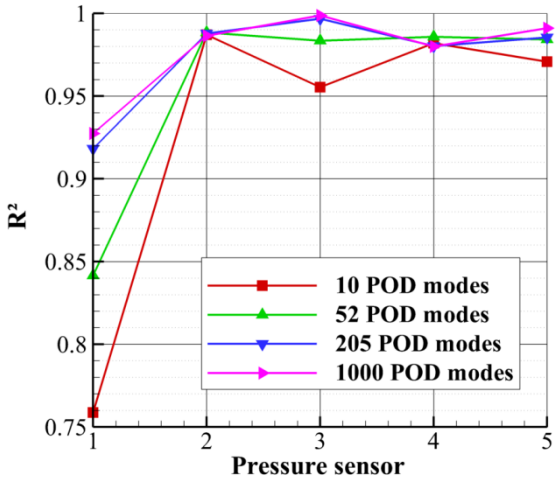


Fig. VI.19: Determination coefficients of "Optim P 5" pressure data estimated by LSE using 10, 52, 205 and 1000 POD coefficients as measurements (validation set, observation model)

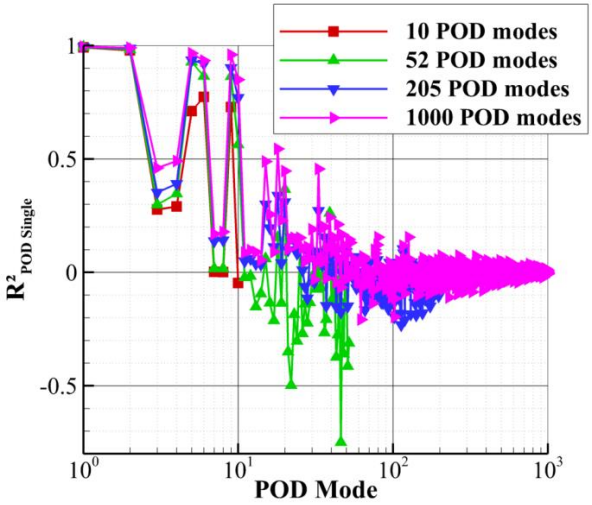


Fig. VI.20: Determination coefficients of POD coefficients estimated using standard Kalman filter and "Optim P 3" set of pressure sensor (validation set)

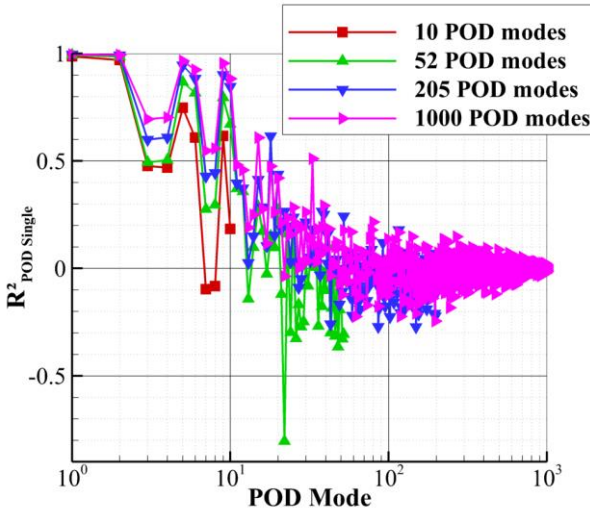


Fig. VI.21: Determination coefficients of POD coefficients estimated using standard Kalman filter and "Optim P 5" set of pressure sensor (validation set)

Standard KF results are displayed in Fig. VI.20 and Fig. VI.21. In these figures, $R^2_{\text{POD Single}}$ of the POD modes estimated by standard KF using "Optim P 3" and "Optim P 5" sets of pressure sensors respectively are plotted. It appears from these figures that in both cases, the best prediction is achieved using 1000 POD coefficients to form the system state. Standard KF results are deteriorated when less POD coefficients are used. Using 205 POD coefficients leads to similar results as the 1000 POD coefficients case, but using 52 or 10 POD coefficients clearly deteriorate the KF performances. However, even with 205 or 1000 POD coefficients, the prediction of POD modes of ranks higher than 40 is poorly accurate ($R^2_{\text{POD Single}} < 20\%$). These observations are slightly surprising, while the prediction of high rank POD modes is not accurate, it seems that it still impacts positively the prediction of low rank POD modes. Also, if these figures are compared with those

obtained by LSE (see Fig. III.23), it appears that the trends remain the same concerning which modes POD are the best estimated among the first 10 POD modes. Both, Kalman filter and LSE have here the same difficulty to predict with high accuracy the POD modes that contains the high frequency content of the flow. That is not so surprising since dynamic and observation models are formed using LSE.

Determination coefficients calculated using the entire velocity field as reference are shown in Table VI.8 where standard KF and SE results are compared. The results confirm previous observation and the best results are obtained when 1000 POD coefficients are used. Values also demonstrate that the standard KF outperforms the LSE with an increase in R^2 of 5 to 10% when only 10 POD coefficients are used and almost 20% with 1000 POD coefficients. The standard KF leads to equivalent and better results than the QSE in the four cases. And even better results than MTD-LSE is obtained with 1000 POD coefficients. These results are consistent with the observations made in the OAT15A URANS case and show once more that more accurate prediction can be achieved with the standard KF than with SE. In addition, we stress that these results are obtained on the validation set and thus are not the results of an overfitting on the training set.

Type of estimation		"Optim P 3"	"Optim P 5"
Standard Kalman filter	10 POD modes	59.9	61.4
	52 POD modes	60.9	65.0
	205 POD modes	64.2	69.3
	1000 POD modes	67.2	72.2
LSE		49.8	54.2
QSE		55.6	62.6
MTD-LSE		67	69.5

Table VI.8: Determination coefficients calculated with the entire velocity field as reference (in %) comparison between standard Kalman filter and SE

4.2.2. Ensemble Kalman filter

4.2.2.1. Quadratic dynamic model and linear observation model

We now look at replacing the linear dynamic model by a quadratic one using the EnKF. It was observed when applying EnKF with the OAT15A URANS data that QSE dynamic models were not always more accurate than LSE dynamic model, especially when the number of POD coefficients used for the system state was big. In such situations, the EnKF with this quadratic model and a linear observation model lead to less accurate predictions than the standard KF. Using OAT15A ZDES database, more POD coefficients are considered to form the system state and thus the QSE model is more likely to be overfitted. That is indeed the case for all four cases. In addition, in the 1000 POD coefficients case, the QSE problem could not be solved since it required the inversion of a matrix or more than 2 To. Only in the 10 POD coefficients case, a QSE dynamic model more accurate than the LSE one is obtained. $R^2_{\text{POD Single}}$ of the best QSE dynamic model (Tikhonov regularization parameter is chosen by CV) are displayed in Fig. VI.22 and Fig. VI.23 for the 10 and 52 POD coefficients state cases. In these figures, the results obtained with a linear dynamic model from LSE are also plotted. In the 10 POD coefficients case, the QSE dynamic model proves to be superior than the LSE model, but it is clear that, in the 52 POD coefficients case, the QSE dynamic model performances are much lower than the LSE dynamic model ones. The same is observed when 205 POD coefficients are utilized (not shown here for brevity).

The EnKF results for the prediction of the velocity field are summarized in Table VI.9 (only the "Optim P 3" set of sensors is considered). As expected, the results are better only in the 10 POD

coefficients case. R^2 is increased by about 5% by using the QSE dynamic model instead of the LSE one in this case. This is approximately the accuracy obtained with the linear dynamic model and 52 POD coefficients. It is however less than the accuracy achieved by the standard Kalman filter with 1000 POD coefficients. On the contrary, the EnKF precision is lower for the 52 and 205 POD coefficients case with QSE dynamic model. The EnKF in this two cases leads to a similar situation as what was observed in the OAT15A URANS with EnKF, QSE dynamic model and 100 POD coefficients. In order to get convergence of the EnKF, QSE dynamic model obtained with a higher Tikhonov regularization parameter than the one given by CV has to be used. It means that even less accurate models than the one shown in Fig. VI.23 have to be employed. Therefore, in the EnKF algorithm, the dynamic model is rapidly disregarded compared with the precision of the observation model and results very close to the LSE are obtained (see annex F.3).

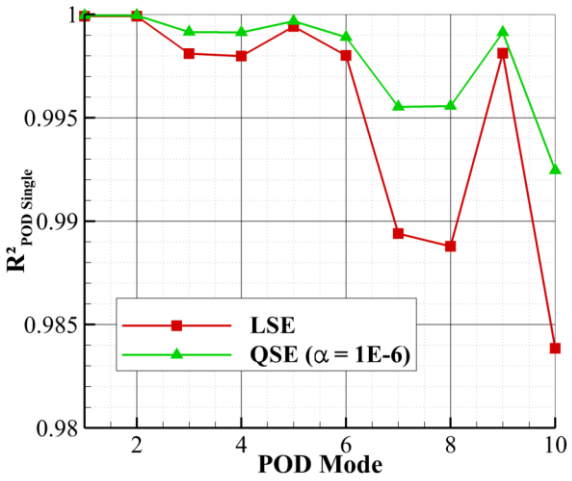


Fig. VI.22: Determination coefficients of the first 10 POD coefficients estimated by QSE and LSE using the same POD coefficients at the previous time step (validation set)

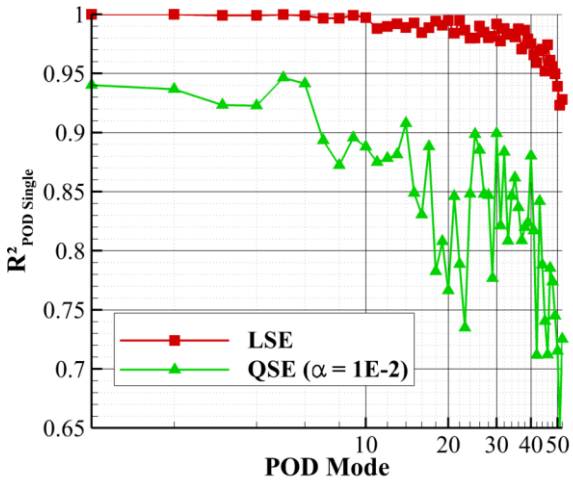


Fig. VI.23: Determination coefficients of the first 52 POD coefficients estimated by QSE and LSE using the same POD coefficients at the previous time step (validation set)

Type of estimation		R^2 (%)
EnKF	10 POD modes	64
	52 POD modes	49.2
	205 POD modes	48.1
	1000 POD modes	-

Table VI.9: Determination coefficients for the prediction of the entire velocity field using EnKF with quadratic dynamic models from QSE, linear observation model and the sensor set “Optim P 3”

4.2.2.2. Linear dynamic model and quadratic observation model

As in the OAT15A URANS case, the impact of using a quadratic observation model is investigated. Both “direct” and “inverse” models are studied. Especially, in the case where 1000 POD coefficients are used, only the “inverse” quadratic observation model is available, since the “direct” model requires the inversion of too large matrices. The results of the EnKF in the several tested situations are shown in Table VI.10. For most cases, the EnKF prediction accuracy is greater using the “inverse” observation model instead of the “direct” one. That is consistent with the observation

made in the OAT15A URANS test case. In addition, “inverse” observation models do not require regularization in the present case, contrary to the direct ones. However, contrary to the OAT15A URANS case, a decrease of R^2 is observed when the number of POD coefficients becomes too large. In particular, it was not possible to obtain a converging EnKF using 1000 POD coefficients. An intermediate case was therefore explored using 500 POD coefficients. The EnKF results with 500 POD coefficients are less precise than those with 205 POD coefficients for “Optim P 3” set and not particularly improved for the “Optim P 5” set of sensors. This is due to the fact that the prediction of high rank POD coefficients is actually very inaccurate, even with QSE. Therefore, taking them into account eventually deteriorates the overall prediction performance of the EnKF. The new information added when using more POD coefficients can for high rank POD coefficients act like a parasitic information. $R^2_{\text{POD Single}}$ for the prediction of the first 1000 POD modes using QSE with the “Optim P 3” set of pressure sensors are displayed in Fig. VI.24. This figure shows that POD coefficients of ranks higher than 50 are not correctly predicted by the “inverse” observation model.

Type of estimation		“Optim P 3”		“Optim P 5”	
		“Direct”	“Inverse”	“Direct”	“Inverse”
EnKF	10 POD modes	59.7	58.9	54.7	64.2
	52 POD modes	57.9	62.5	62.8	70.1
	205 POD modes	50.3	73.2	50.7	79.4
	500 POD modes	-	69.9	-	79.9
	1000 POD modes	-	Diverging	-	Diverging

Table VI.10: Determination coefficient (in %) of the prediction of the entire velocity field using EnKF with linear dynamic model and quadratic observation model

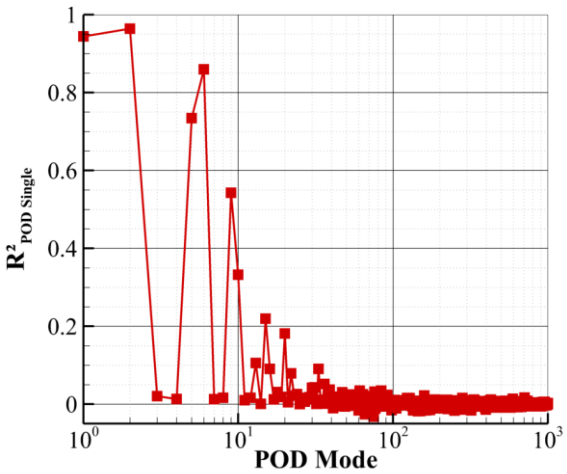


Fig. VI.24: Determination coefficients of the first 1000 POD coefficients estimated by QSE with the “Optim P 3” set of sensors (validation set)

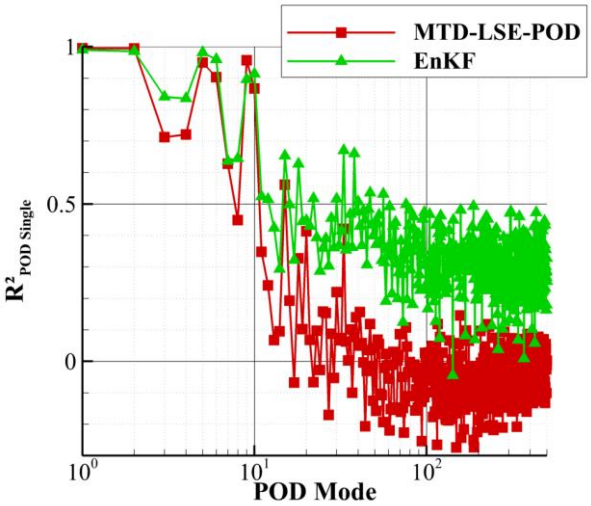


Fig. VI.25: Determination coefficients of the first 500 POD coefficients estimated by MTD-LSE-POD and EnKF (linear dynamic model and “inverse” quadratic observation model) with “Optim P 5” set of sensors (validation set)

Comparing these results with those of the standard KF (in which only linear models are employed), it appears that the prediction ability of the EnKF with “inverse” quadratic observation model is superior to the standard KF one. R^2 of the standard KF prediction of the velocity field using 500 POD coefficients are 66.5% and 72.3% for “Optim P 3” and “Optim P 5” set respectively. This

confirms the positive impact of using “inverse” quadratic observation model on the EnKF prediction performances. But it also highlights the negative impact that using too many POD coefficients can have on the EnKF performances. In fact, looking at the standard KF results and considering the 500 POD coefficients case, it is already possible to see, when “Optim P 5” set of sensors is used, that the EnKF performances are not really improved between the 500 and 1000 POD coefficients cases.

At last $R^2_{\text{POD Single}}$ for the first 500 POD modes predicted by the EnKF with linear dynamic model and “inverse” quadratic observation model (which corresponds to the best prediction obtained with EnKF so far) are compared with those predicted by MTD-LSE-POD (best prediction obtained by SE methods) in Fig. VI.25. This figure shows that EnKF outperforms the MTD-LSE-POD. In particular, for POD modes of ranks higher than 40, the MTD-LSE-POD prediction is very inaccurate ($R^2_{\text{POD Single}}$ around 0 and negative) while the EnKF results are far better with an average $R^2_{\text{POD Single}}$ about 30% up to POD mode 500.

4.2.2.3. Quadratic dynamic and observation model

From the previous observations made for the use of quadratic observation models, it is decided to use only “inverse” quadratic observation models. As observed in the OAT15A URANS case, the “inverse” quadratic observation model allows a better stability of the EnKF with quadratic dynamic models. As such, no regularization is necessary for the dynamic model with 10 POD coefficients and lower Tikhonov regularization parameters are required to obtain the best results.

Determination coefficients of the prediction of the entire velocity field are given in Table VI.11. Compared to R^2 values of the EnKF prediction using linear dynamic model and “inverse” quadratic model, R^2 is increased only in the 10 POD coefficients case. In the 52 POD coefficients case, the EnKF accuracy is similar to the one of the EnKF with linear dynamic model. For the 205 POD coefficients case, the EnKF is deteriorated. That is perfectly consistent with the observations made when using EnKF with quadratic dynamic model and linear observation model. It was shown that only in the 10 POD coefficients case a QSE dynamic model more accurate than the LSE dynamic model was obtained. Therefore, in this situation the EnKF performs better with quadratic dynamic model than linear one whatever the observation model used. On the contrary, QSE dynamic models formed in the 52 and 205 POD coefficients case were less accurate than the LSE ones, even when using the best regularization parameter obtained by CV. Therefore, there is no reason why their use should improve the EnKF performances. That is what is observed.

Type of estimation		“Optim P 3”	“Optim P 5”
EnKF	10 POD modes	65.2	68.4
	52 POD modes	61.6	70.1
	205 POD modes	50.9	61.8
	1000 POD modes	-	-

Table VI.11: Determination coefficients for the prediction of the entire velocity field using EnKF with QSE dynamic and observation models

4.2.2.4. Radial Basis Function dynamic model

It was previously explained that quadratic dynamic model could not be formed using QSE and 1000 POD coefficients because it required the inversion of a too large matrix. In these circumstances, the use of RBF could allow to form a nonlinear model of the dynamic with so many coefficients as system state.

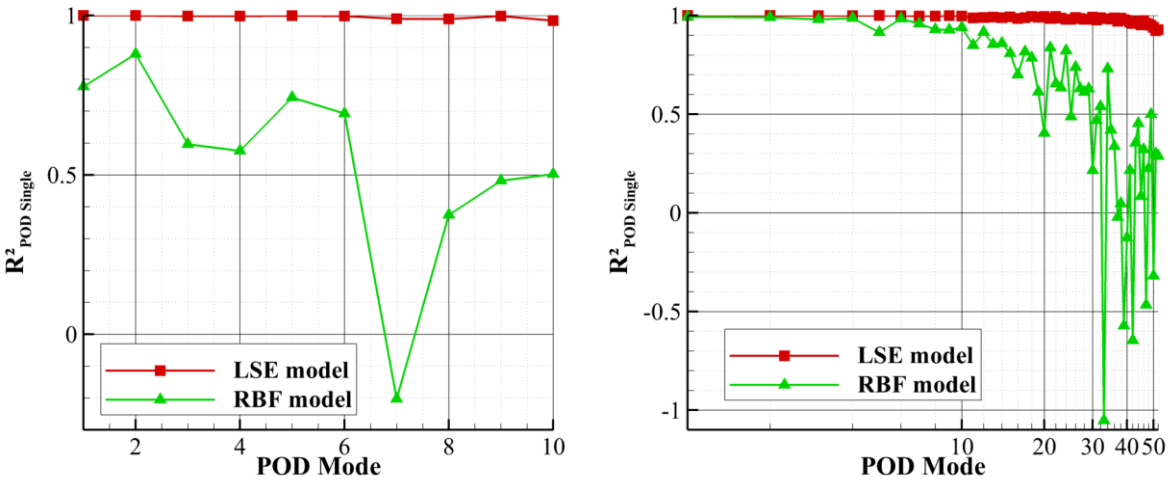


Fig. VI.26: Determination coefficients of POD coefficients estimated by LSE and RBF network using 10 and 52 POD coefficients at the previous time step as measurements (validation set; a) 10 POD coefficients, b) 52 POD coefficients)

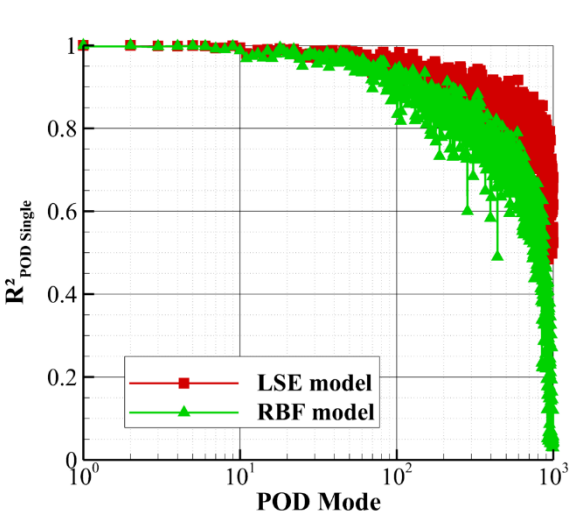


Fig. VI.27: Determination coefficients of POD coefficients estimated by LSE and RBF network using 1000 POD coefficients

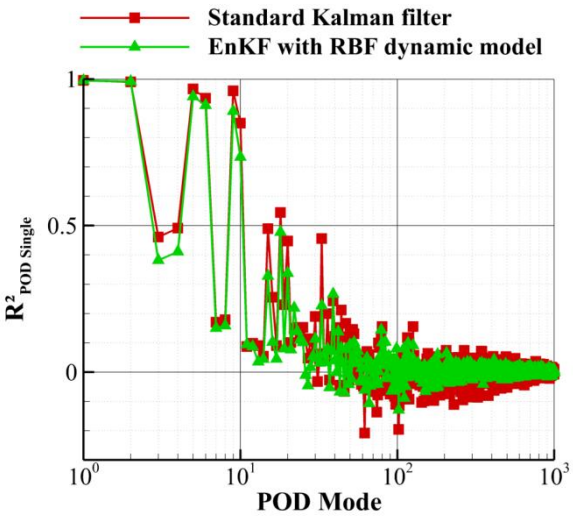


Fig. VI.28: Comparison of determination coefficients of POD coefficients estimated using standard Kalman filter and ensemble Kalm filter with RBF dynamic model (validation set; 1000 POD coefficients; 3 pressure sensors)

RBF dynamic models were formed in the 10, 52 POD and 1000 POD coefficient state cases. The same procedure as in the OAT15A URANS case is employed to create the RBF dynamic models. $R^2_{\text{POD Single}}$ calculated for the estimation (by RBF and LSE dynamic models) of the system state POD coefficients on the validation set are plotted in Fig. VI.26 and Fig. VI.27. The figures clearly demonstrate that the RBF dynamic models are less accurate than the LSE ones. As demonstrated previously such dynamic models will not improve the EnKF performance when used instead of LSE dynamic models. The EnKF was nevertheless run using the RBF dynamic model with 1000 POD coefficients and 3 pressure sensors. $R^2_{\text{POD Single}}$ of the prediction of these POD coefficients are displayed in Fig. VI.28. The EnKF, with this RBF dynamic model, is less accurate than the standard

KF. That is confirmed by the value of R^2 computed for the estimation of the entire flow field which reaches 65.3% instead of 67.2% with the standard KF.

4.3. Wall mounted cube

Kalman filter is now applied to the estimation of POD coefficients of the wall mounted cube flow. The goal is now to verify if the previous observations still hold for an experimental case for which the dynamic is more complex. In addition, it was shown previously that LSE does not perform well in this test case. Therefore, it is interesting to assess how an improvement will the Kalman filter be compared to SE methods, but also to check if it is enough to generate dynamic and observation models using LSE and QSE in such a case.

Only the data from the high speed PIV campaign are used since it is necessary to have time correlated information in order to form the dynamic model. For brevity, only the vertical plane, for the 10 and 30 $m.s^{-1}$ cases, is studied. In both cases, 3 POD coefficients state are considered since it was shown that the first 3 POD coefficients are far better estimated than all the others (see Chapter III.4.2). Two other states are used for which the number of POD coefficients corresponds to 70 and 90% of TKE (respectively 58 and 333 POD coefficients in the 10 $m.s^{-1}$ case; 71 and 360 POD coefficients in the 30 $m.s^{-1}$ case). At last a 1000 POD coefficients case is also investigated. The conditional events are the 23 wall pressure measurements.

4.3.1. Standard Kalman filter

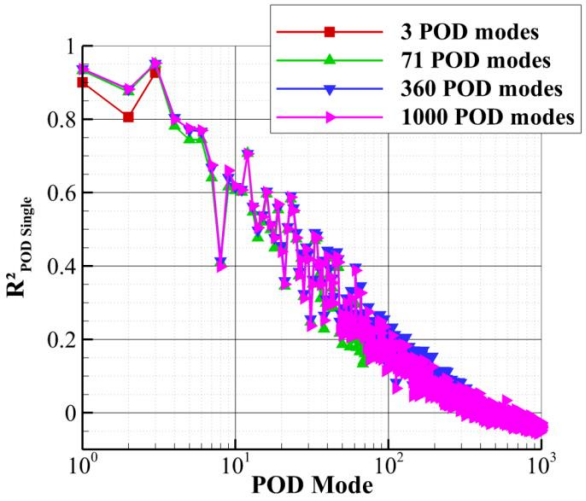


Fig. VI.29: Determination coefficients of POD coefficients estimated by LSE using 3, 71, 360 and 1000 POD coefficients at the previous time step as measurements (validation set, dynamic model)

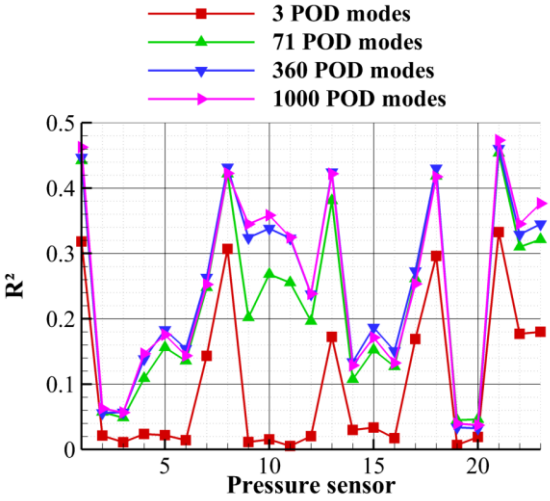


Fig. VI.30: Determination coefficients of the pressure data estimated by LSE using 3, 71, 360 and 1000 POD coefficients as measurements (validation set, observation model)

LSE dynamic models accuracy is once more evaluated through the calculation of $R^2_{\text{POD Single}}$ for the prediction of the system state at time $t+\Delta t$ by LSE using the system state at time t as measurements. Values are plotted in Fig. VI.29. Contrary to OAT15A cases, only the first 6 POD modes are accurately predicted by the LSE dynamic model. In addition, no real improvement is observed on the prediction accuracy between the 71, 360 and 1000 POD modes cases. At last, from about POD mode 300, the LSE dynamic model is very inaccurate with values of $R^2_{\text{POD Single}}$ going negative.

LSE observation models accuracy is shown in Fig. VI.30, where the R^2 of the prediction of the pressure sensors using the POD coefficients as measurements are plotted. The impact of increasing the system state size is more important on the observation model performance than on the dynamic model one. Adding POD coefficients improves the estimation of the pressure data, however there seems to be a threshold as 360 POD coefficients and 1000 POD coefficients cases give very similar results. Also, all the observation models are far less accurate than those obtained in OAT15A cases. For all sensors R^2 is less than 50% and for most it is less than 30% whereas in the OAT15A ZDES case R^2 was always higher than 75%. In overall, both linear dynamic model and linear observation model are mildly accurate in the present test case. It is coherent with SE results shown in Chapter III.2.

Type of estimation		R^2 (%)
Standard Kalman filter	3 POD modes	15.3
	71 POD modes	20.2
	360 POD modes	20.6
	1000 POD modes	20.8
LSE		13.4
QSE		14.6
MTD-LSE		22.4

Table VI.12: Determination coefficients (in %) calculated with the entire velocity field as reference comparison between standard Kalman filter and SE (30 m.s^{-1} case)

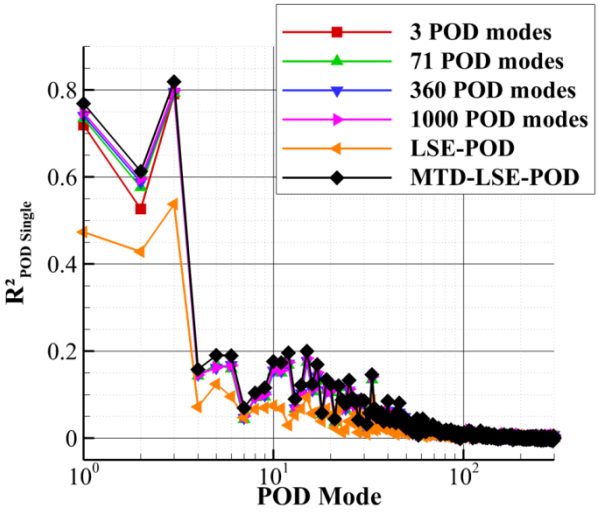


Fig. VI.31: Determination coefficient of POD coefficients estimated using standard Kalman filter with system state of 3, 71, 360 and 1000 POD coefficients (validation set). Results of the LSE-POD and of the best MTD-LSE-POD situation are shown for comparison.

Standard KF results are summarized in Table VI.12 where values of R^2 computed for the prediction of the entire velocity field are reported. The results obtained with the standard KF are also better than the LSE and QSE ones, whatever the number of POD coefficients considered to form the system state. That agrees with observations made in the OAT15A cases. However, contrary to the results obtained in OAT15A test cases, standard KF results while close are not better than the MTD-LSE results. As in OAT15A cases, increasing the system state size improves the standard KF results. Nevertheless, differences are almost negligible between 71, 360 and 1000 POD coefficients cases. Looking at the prediction accuracy of each POD coefficient separately (see Fig. VI.31), it seems that increasing the number of POD coefficients has still a positive impact on the

prediction accuracy of the first 3 POD coefficients. Thus, the improvement observed on R^2 for the entire velocity field does not simply come from the fact that more POD modes are considered. Comparing these results with those of the MTD-LSE-POD, it is seen that MTD-LSE-POD actually outperforms the standard KF for all POD modes in this test case.

Results obtained for the vertical plane at 10 $m.s^{-1}$ are shown in Table VI.13 and are perfectly consistent with those obtained at 30 $m.s^{-1}$. Observations concerning the LSE dynamic model and LSE observation model are also similar. Therefore, in the following only the 30 $m.s^{-1}$ case will be presented. In addition, the 1000 POD coefficients case is disregarded.

Type of estimation		R^2 (%)
Standard Kalman filter	3 POD modes	19.3
	58 POD modes	23.5
	333 POD modes	25.2
LSE		20.4
QSE		23.3
MTD-LSE		27.8

Table VI.13: Determination coefficients (in %) calculated with the entire velocity field as reference comparison between standard Kalman filter and SE (10 $m.s^{-1}$ case)

4.3.2. Ensemble Kalman filter

4.3.2.1. Quadratic dynamic model and linear observation model

Tests using OAT15A URANS and ZDES database shown that, most of the time, the use of QSE dynamic model does not improve the KF prediction performances, especially when a large number of POD coefficients is used to form the system state. Indeed, due to overfitting, Tikhonov regularization becomes necessary and leads to a dynamic model that is eventually less accurate than the linear one obtained by LSE. Therefore, in the wall mounted cube case, except in the 3 POD coefficients case, no QSE model is more accurate than the LSE one and no improvement is expected. In addition, for the 360 POD modes case, no QSE dynamic model was generated due to Random Access Memory storage limitations.

In the 3 POD coefficients case, the QSE dynamic model is slightly more accurate than the LSE one without requiring Tikhonov regularization. However, $R^2_{\text{POD Single}}$ is increased by less than 1% for each POD coefficients. Finally, this improvement does not lead to more accurate EnKF results compared to the standard KF (or to the EnKF using linear dynamic model for which R^2 is 15%). On the contrary, a slight deterioration is observed in term of R^2 for the entire flow field (see Table VI.14). Using 71 POD coefficients, Tikhonov regularization is necessary to compensate the overfitting when generating the QSE dynamic model. The QSE dynamic model obtained is finally less accurate than the LSE dynamic model. The EnKF results using this QSE dynamic model are also less accurate than those of the standard KF (see Table VI.14).

Type of estimation		R^2 (%)
EnKF	3 POD modes	14.9
	71 POD modes	13
	360 POD modes	-

Table VI.14: Determination coefficients for the prediction of the entire velocity field using EnKF with quadratic dynamic models from QSE and linear observation model (vertical plane, 10 $m.s^{-1}$)

4.3.2.2. Linear dynamic model and quadratic observation model

In the OAT15A cases, the use of quadratic observation model, and especially "inverse" quadratic observation model, showed, most of the time, clear improvements. In the wall mounted test case, "direct" quadratic observation models are employed only with 3 and 71 POD coefficients. In the present test case, results of using quadratic observation models are much more mitigated. The determination coefficient of the prediction of the velocity field by EnKF with linear dynamic model and quadratic observation model are given in Table VI.15. None of the tested situations result in a better prediction than the standard KF. On the contrary, the prediction accuracy is slightly deteriorated. That may come as a surprise in comparison with the results obtained in the OAT15A cases, but it simply comes from the fact that the "direct" quadratic observation models are actually not more accurate than the linear ones. Fig. VI.32 shows determination coefficients of the prediction of the 23 pressure sensors separately using LSE and QSE (with Tikhonov regularization, best regularization parameter chosen by CV) and the 71 first POD coefficients as conditional events. This figure demonstrates that the "direct" QSE observation model does not yield any gain compared to the "direct" LSE observation model.

Type of estimation		"Direct"	"Inverse"
EnKF	3 POD modes	15.2	14.8
	71 POD modes	19.6	18.8
	360 POD modes	-	18.9

Table VI.15: Determination coefficient (in %) of the prediction of the entire velocity field using EnKF with linear dynamic model and quadratic observation model

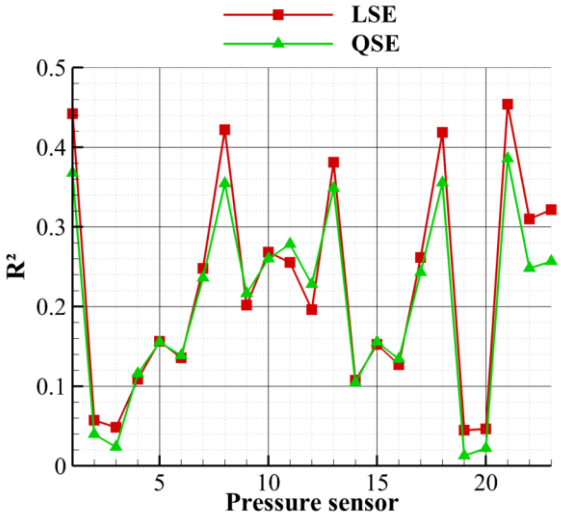


Fig. VI.32: Determination coefficients of the pressure data estimated by LSE and QSE using 71POD coefficients as measurements (validation set)

For the "inverse" observation models, the deterioration can already be observed for an "inverse" linear observation model. And compared to the "inverse" linear observation model, the "inverse" quadratic observation model is an amelioration and lead to a more accurate EnKF. Indeed, R^2 of the EnKF using "inverse" linear observation model is about 17%. That is 3% less than if a "direct" LSE observation model is used and 2% less than when "inverse" QSE observation model is used. The fact that the "inverse" QSE observation model forms a better EnKF than the "inverse" LSE one is consistent with the fact that it is a more accurate observation model (which was observed in Chapter III.4.2 where QSE was shown to better perform than LSE). But it is unclear why the

"inverse" linear observation model deteriorates the EnKF results. Indeed, both "direct" and "inverse" models do not suffer from strong overfitting. They are nevertheless two different models, with different error covariance matrices and thus can lead to different results. In addition, deterioration of the EnKF prediction accuracy when using "inverse" linear observation models was also observed in the OAT15A ZDES test case (see Table VI.10).

4.3.2.3. Quadratic dynamic and observation model

From the previous results using QSE dynamic model and QSE observation model separately, no amelioration of the EnKF accuracy should be expected by combining the two. That is quickly assessed using 71 POD coefficients. Using "direct" QSE observation model, a determination coefficient for the prediction of the velocity of 20% is reached. This R^2 is of the same order of magnitude than the one obtained with "direct" QSE observation model and LSE dynamic model or by the standard KF, but not better. When using the "inverse" observation model R^2 goes down to 14.3% which represent an improvement compared to the use of a linear observation model with the "inverse" QSE observation model, but not compared to the use of the "direct" QSE observation model. The deterioration, compared to the "direct" QSE observation model is consistent with the observation made when using QSE observation model with LSE dynamic model previously.

Those results obtained using QSE models show the limitations of generating quadratic dynamic or observation models by QSE. In these conditions, RBF model is once more investigated to see if it may provide a better alternative to form nonlinear dynamic models than SE.

4.3.2.4. Radial Basis Function dynamic model

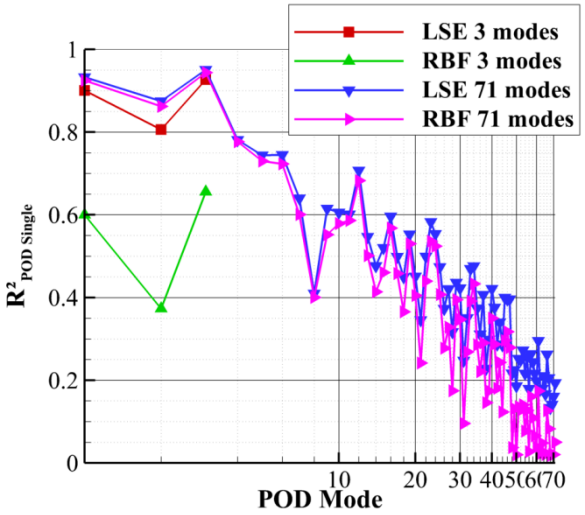


Fig. VI.33: $R^2_{\text{POD Single}}$ of POD coefficients estimated by LSE and RBF network using 3 and 71 POD coefficients at the previous time step as measurements (validation set)

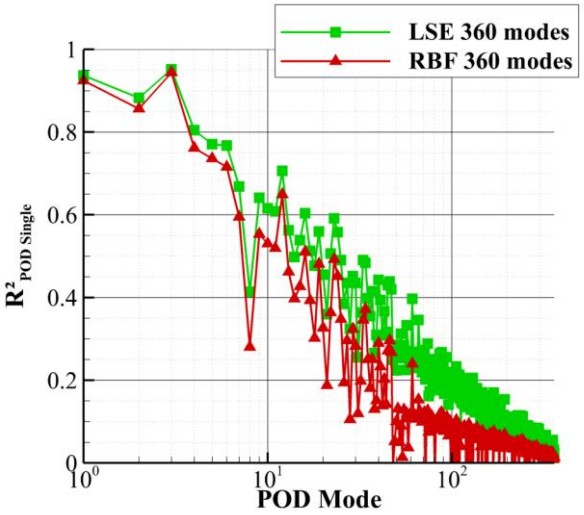


Fig. VI.34: $R^2_{\text{POD Single}}$ of POD coefficients estimated by LSE and RBF network using 360 POD coefficients at the previous time step as measurements (validation set)

RBF dynamic models are generated for the 3, 71 and 360 POD coefficients cases. 18420 time samples from the training are used as centers for the RBF and its prediction accuracy is evaluated on the remaining 6140 samples of the validation set. The results are displayed in Fig. VI.33 and Fig. VI.34. In all situations, RBF models lead to less accurate prediction for all POD modes compared with LSE models. Especially, in the case where only 3 POD coefficients are considered, the deterioration is particularly strong. For the other two cases, the deterioration is smaller, and the RBF model results are close to the LSE model ones, but seem to be more deteriorated for high rank POD modes. As demonstrated earlier, such dynamic RBF models will not bring any improvement if use

with the EnKF instead of the LSE dynamic models. They are not a correct alternative to the QSE dynamic models. One thing that could be done to improve the RBF models and was not experimented, is to choose different shape parameters for all the RBF in the sum for one POD coefficient (see Chapter VI.3.5.1). Using other functions than Gaussian one is also another possibility.

5. Chapter summary

The use of Kalman filter (KF) has been investigated using the OAT15A URANS and ZDES databases, as well as the wall mounted cube high speed PIV database, as test cases. First, the creation of the dynamic and observation models necessary for the KF was discussed. For dynamic model, a popular possibility is to use POD Galerkin projection of the Navier-Stokes equations. It is then possible to show that the time-derivative of one POD coefficient is the sum of three terms: a linear function of the other POD coefficients corresponding to the dissipation, a quadratic functions of the other POD coefficients corresponding to the convection and a constant term from the pressure contribution. Therefore, when generating a dynamic model for the POD coefficients, one is looking at a quadratic function on the POD coefficients. Computing directly the POD Galerkin model can be challenging. Most of the time, using experimental data, the pressure term cannot be directly evaluated and the resulting model needs a careful tuning. Therefore, alternative methods were investigated: the LSE, the QSE, the RBF network and 4D-variational data assimilation. The same methods were also employed to generate observation models.

In the OAT15A URANS and ZDES test case, the standard KF using LSE dynamic and observation models was shown to outperform the LSE, and even the QSE and MTD-LSE when enough POD coefficients were taken into account in the system state. For these test cases, LSE dynamic and observation models were shown to be highly accurate and taking into account more POD coefficients in the system state improves the prediction. Extended KF (EKF) was also studied and its results were not better than those of the standard KF. EKF proved to be difficult to converge and required a careful tuning of the QSE dynamic model. The QSE dynamic model had to be generated using Tikhonov regularization with too high regularization parameters resulting in a dynamic model that was not particularly more accurate than the LSE one. Data assimilation was explored to form a quadratic dynamic model from the LSE one. The EKF results using this model were similar the standard KF.

The Ensemble KF (EnKF) was thus used as an alternative to the EKF in order to use quadratic dynamic and observation models. Using QSE dynamic model with LSE observation model, it was possible to improve the prediction accuracy of the EnKF compared to the standard KF, but not in all cases. When too many POD coefficients were taken into account in the system state, the QSE dynamic model became strongly overfitted and lead to a deterioration of the prediction compared to LSE model. The difficulty to obtain a good QSE dynamic model was especially demonstrated in the OAT15A ZDES case. In this case, performance amelioration was obtained only when 10 POD coefficients were taken into account.

The use of QSE observation model was shown to hold some strong improvements but mostly if an “inverse” observation model was used. Using QSE observation models appeared to be a better way to improve the KF performances in the OAT15A test cases than using QSE dynamic model. At last, the use of both QSE dynamic and “inverse” observation models held some improvements if the QSE dynamic model was not strongly overfitted. In the OAT15A URANS test case, the best prediction was obtained with QSE dynamic model and QSE “inverse” observation model using 50 POD coefficients. In the OAT15A ZDES case, the best prediction was made using LSE dynamic model and QSE “inverse” observation model with 500 POD coefficients in this case. R^2 of the entire velocity field was increased by 10% compared to the best MTD-LSE. The use of RBF network to

form nonlinear dynamic model was also investigated. In both URANS and ZDES case, RBF network formed less accurate dynamic model than LSE. The EnKF results using the RBF models were not better than those obtained with LSE.

Test of the KF on the OAT15A URANS and ZDES cases showed that its accuracy was higher than LSE and QSE most of the time, and that EnKF with QSE observation model could also outperform the MTD-LSE. Both standard KF and EnKF were thus applied to the wall mounted cube high speed PIV database.

The results obtained in this last application are less encouraging than those obtained with the OAT15A cases. The standard KF was still more accurate than the LSE and QSE when enough POD modes formed the system state. But it was not more accurate than the MTD-LSE. With the EnKF, the use of QSE dynamic model lead to lower performances of the filter and it was not possible to generate by QSE a more accurate dynamic model than by LSE (except when 3 POD coefficients were taken into account). The same problem occurred with QSE observation models and no amelioration were obtained using QSE “direct” or “inverse” observation models. RBF dynamic models were also shown to remain less accurate than LSE ones. In the wall mounted test case, the best KF results were thus obtained by the standard KF with LSE dynamic and observation models and these results were less accurate than those of the MTD-LSE. Also the best dynamic models were obtained by LSE but they actually form poorly accurate models. In the wall mounted cube test case, the flow dynamic is clearly 3D (which was not the case in the OAT15A case) therefore it is likely that by considering only one plane the dynamic could not be captured.

Conclusions and perspectives

This work lies in the context of real-time flow monitoring. In particular, the main objective of this study was to investigate methods for the reconstruction and especially the prediction of turbulent flows from a limited number of sensors.

The literature review, focused on the Stochastic Estimation (SE), shwoed that the SE is a simple and very general method that leaves the user with a great degree of freedom concerning the data to estimate and the sensors used to make the estimation. The past applications of SE demonstrated that it was a promising method to predict turbulent flow states. However, few studies were found to be devoted to the quantification of the SE accuracy to provide real-time prediction of turbulent flows. In addition, the few that were mainly concerned low Reynolds turbulent flows, in which most of the Turbulent Kinetic Energy could be retained by a few POD modes. It was therefore decided to first study the applications of SE on increasingly complex flows and then to better characterize the methods in relation to the turbulent content of the flow.

The theoretical background was described in Chapter II. In particular, the error criteria by which the estimation methods would be compared were detailed. Special attention was also given to the convergence of the statistical moments necessary to the training of the SE.

In Chapter III, SE and most of its extensions (Linear Stochastic Estimation LSE, Quadratic SE QSE, Multi-Time-Delay Stochastic Estimation MTD-SE, SE combined with Proper-Orthogonal-Decomposition POD) were applied to four flow configurations. Their performances were evaluated and compared. The LSE was shown to hold very accurate prediction (estimation of data outside the training set) of the simulated wake flow of a blunt trailing edge, as expected from the literature review. The accuracy of the LSE was lower when applied to the prediction of the flow around an OAT15A airfoil simulated by ZDES. In particular, it was observed that the SE methods badly predicted the vortex shedding developing in the wake of the airfoil. The shock oscillation was, nevertheless, well captured. The LSE predictions for the Backward Facing Step (BFS) and wall-mounted cube experimental test cases where, on the contrary, poorly accurate.

LSE and QSE was compared and it was demonstrated that QSE did not systematically hold a more accurate prediction (estimation outside the training set). This was due to the overfitting that QSE is more likely to suffer than LSE and the use of Tikhonov regularization led only to slight improvements of the prediction quality.

For all test cases, the MTD-LSE was found to be the most accurate method. However, these good results were only obtained using Tikhonov regularization. In the four test cases, it was also showed that delays had to be chosen inside an optimal time window. Adding delays from outside this window actually deteriorated the prediction accuracy. Nevertheless, for the BFS and, to a lesser extend, for the cube, the MTD-LSE accuracy remained poor.

At last, the use of LSE-POD and QSE-POD did not particularly improve the reconstruction nor the prediction accuracy. But, it was possible to reduce the overfitting that the QSE suffered from by not considering the highest order POD modes. The main interest in using LSE-POD, is that the estimation problem can be reduced to the estimation of a smaller number of POD coefficients than the number of spatial points in the field to estimate.

Following the poorly accurate results obtained in the BFS and in the wall mounted cube cases, it was decided to better characterized the SE methods (Chapter IV).

First, the impact of Gaussian noise on the data (conditional events and/or data to estimate) was investigated. It was demonstrated that the Gaussian noise contained in the data to estimate was filtered out. Therefore, if noisy data are used to calculate the error, the performance of the SE methods are underestimated. But the main results of this investigation was the emphasis of the strong impact of the noise contained in the conditional events. Such noise directly penalizes the estimation accuracy. Another important result is the demonstration that SE can present a strong overfitting if trained with clean data and used with noisy data.

Then, the impact of the turbulence on the SE method performances is investigated by comparing estimations of the flow around the OAT15A airfoil in transonic conditions, obtained using URANS (turbulence modelized) and ZDES (turbulence partially resolved) simulations. The velocity fields simulated by URANS were predicted with a higher accuracy than the one from the ZDES simulation. The study of the spectral content of the flow highlighted that the low frequency content was better estimated than the high frequency one. These observations were confirmed by studying the impact of temporally filtering the conditional events and/or the data to estimate in the BFS test case.

In addition, it was observed that the resolution of part of the turbulence in the ZDES simulation led to a shortening of the turbulent spatial integral length scales in the wake. Turbulent spatial integral length scale maps were compared with maps of the normalized mean square error. This comparison showed a clear relation between the high error areas of the field and the areas of shortest integral length scales. That indicated that the short integral length scales were estimated with less accuracy than the long ones.

From these observations, the estimation of the turbulent spatial integral length scales by LSE was investigated more precisely on the BFS case (U_0 = 30 m.s^{-1}). The LSE was shown to not conserve the turbulent spatial length scales and to filter the shortest length scales. The study of the POD modes estimation confirmed this conclusion. By associating a characteristic length scale to each POD mode, it was observed that the POD modes estimated with the highest accuracy were also those associated to the longest characteristic lengths. The study of the spatial filtering of the velocity fields was also carried out and confirmed these results.

Using the data from the wall-mounted cube experiments, the impact of the Reynolds number (Re) on the estimation quality of the same flow geometry was studied. This investigation showed that the estimation became more accurate when the Reynolds number was decreased. Looking at the estimation of POD modes, it was observed that the accuracy of their estimations is improved with decreasing Reynolds number. In addition, the number of POD modes required to retain the same level of TKE decreases with the Reynolds number, and more and more energy was contained in the large structures. The POD modes that are the best estimated represents then a larger and larger part of the flow energy content when the Reynolds number decreases which improves the overall estimation accuracy.

Also the possibility to use a SE model to estimate flows for which it had not been trained was investigated. The results highlighted that, to some extent, training the LSE with normalized data allowed to predict flow at different velocity upstream conditions without a strong deterioration of the accuracy. This preliminary study thus demonstrated the possibility to use LSE for a range of upstream conditions of a same flow geometry.

At last, in the wall-mounted cube experiments, it was demonstrated that the LSE accurately predict phase-averaged fields, even if the prediction of instantaneous velocity fields was inaccurate. An extension of the LSE, called “phased”-LSE, in which several sets of LSE coefficients trained for different phases of the flow are used was presented. The predictions obtained were more accurate than the LSE and QSE ones, but not than MTD-LSE ones. The study could not however be performed with many phase segments due to convergence issues. A perspective could be to try the

method with richer databases to better examine the possible improvements on the accuracy that it offers.

The impact of the conditional event locations on the estimation accuracy was also observed in several test cases. In particular, the study of the turbulent integral length scales using POD shown that a POD mode was better estimated when conditional events were located close to its spatial extrema. It was therefore decided to investigate the opportunity to optimize the sensor location algorithm, in order to improve the accuracy of the estimation obtained by SE methods, in Chapter V.

To optimize these locations, an algorithm first proposed by Muradore et al. was examined. The optimized locations, for the estimation of one POD mode, were found to be close to its extrema, which was consistent with observations in the literature and from the previous chapter. This observation explains for a part the difficulty to predict velocity field using only wall measurements and SE methods. The ability of the algorithm to find locations that lead to higher reconstruction and prediction accuracy, using LSE, was verified and confirmed using the OAT15A airfoil and BFS databases. However, with QSE, some optimized sets of sensors performed with less accuracy than intuitively chosen ones. Nevertheless, the algorithm represents a convenient tool to determine the number and the locations of sensors necessary to achieve a chosen level of accuracy on one or several POD modes.

Since MTD-LSE was demonstrated to improve the estimation accuracy, an extension of the sensor location optimization algorithm was proposed to choose both locations and delays. This extended algorithm was first applied to the choice of delays from a limited number of sensor locations chosen a priori. The results highlighted the possibility to improve the prediction accuracy using a small number of delays. In particular, it was possible to reach same levels (but not higher) of determination coefficient than in the best MTD-LSE results, using constant time step delays, with less delays (and Tikhonov regularization).

The possibility to decrease the number of conditional events necessary to reach a chosen level of accuracy on the estimation by using both synchronized and delayed information was also examined. Taking into account delayed information was particularly useful when the space of possible locations was restrained. At last, a flow chart for the utilization of (MTD)-LSE-POD was then proposed.

From the characterization of the SE, it appeared that even using sensors in flow with optimized locations, the prediction accuracy will remain low in the BFS case and in the wall mounted cube case. SE methods were therefore compared with the Kalman filter.

After a short literature review and reminder of the mathematical background, the use of the Kalman filter was investigated using the OAT15A URANS and ZDES databases, as well as the wall mounted cube high speed PIV database (Chapter VI). From the literature overview, it was decided to use POD coefficients to form the system state. To avoid the direct computation of the coefficients of the POD ROM from POD Galerkin projection of the Navier-Stokes equations, models generated by LSE, QSE (potentially corrected by 4D-Var) and Radial Basis Function (RBF) networks were employed.

In the OAT15A URANS and ZDES cases, the standard KF using LSE dynamic and observation model was shown to outperform the LSE. When enough POD coefficients were taken into account in the system state, even the QSE and MTD-LSE were outperformed. In such situations, the dynamic and observation models were highly accurate and taking into account more POD coefficients improved the prediction accuracy. Extended KF (EKF) was also applied, but it proved to be difficult to converge and required a careful tuning of the dynamic model obtained by QSE. The Ensemble KF (EnKF) was then used as a more robust alternative in which quadratic dynamic model

could be employed, as well as quadratic observation models. If it was possible to get higher accuracy with the EnKF (and QSE dynamic model) than with standard KF in some situations, when too many POD coefficients were taken into account, the QSE dynamic model became overfitted and deteriorated the prediction accuracy, even using Tikhonov regularization. The use of QSE observation model hold strong improvements but mostly when the "inverse" observation model was utilized. Using QSE observation models appeared to be a better way to improve the KF performances, in the OAT15A cases, than using QSE dynamic models. Eventually, the best results in the OAT15A URANS case were obtained by the EnKF with a QSE dynamic model and a QSE "inverse" observation model. But for the OAT15A ZDES case, the best results were from the EnKF with LSE dynamic model and QSE "inverse" observation model. In this second case, R^2 was increased by 10% compared to the best MTD-LSE results. For both OAT15A URANS and ZDES simulations, RBF network formed less accurate dynamic model and did not lead to more accurate EnKF than LSE or QSE models.

The results of the applications of KF on the wall-mounted cube high speed PIV database (U_0 = 30 m.s^{-1}) are somehow less encouraging than the results from the OAT15A test cases. Standard KF was still found more accurate than LSE and QSE, but not than MTD-LSE. Using EnKF, no QSE dynamic or observation models lead to an accuracy improvement compared to the standard KF. The study showed the importance of the dynamic and observation models quality. The poor performances of the EnKF (and standard KF) are thus related to the use of LSE and QSE to form the models. In this last test case, LSE and QSE seemed inappropriate to generate these models. In addition, it may be due to the fact that only one plane was considered, which is probably not enough to capture the dynamic of the flow that is three dimensional.

In overall, this study has particularly highlighted the limitations of SE to predict with high fidelity the velocity content of highly turbulent flows. The smallest turbulent spatial integral length scales cannot be predicted with a limited number of sensors. When the ranges of turbulent scales and of frequencies increase, SE methods seem to become less accurate. As Durgesh et al. stated: "the multi-time-delay LSE approach should successfully capture the dynamics in any turbulent flow that has significant energy in a few POD modes". The results presented in this work demonstrates that the MTD-LSE does not capture the dynamics of turbulent flows that require a large number of POD modes to capture a significant amount of energy. Nevertheless, SE methods may be enough for some flow control applications, even for highly turbulent flows. As proposed in the flow chart in Chapter V.3, if the control requires only the real-time prediction of a few identified POD modes, then SE may be enough to obtain such prediction with a sufficient accuracy using optimized sensor positions. Others flow decomposition methods could also be considered without modifiying the flow chart.

In the case where such conditions could not be reached, other techniques should be preferred. In particular, methods exploiting the dynamic and physic knowledge of the flow can be of great interest. For instance, methods based on model of the flow. A downside of this kind of methods is that they are highly dependant of the applications and that deriving the model can be tricky and expensive. But such methods can clearly have better prediction performances than SE ones. This was demonstrated using the Kalman filter with very simple models. One can fairly expect the Kalman filter to perform much better with more complex models of the system. It would be interesting to try the Kalman filter using dynamic or observation models obtained through other means than linear regression. Other decomposition methods such as the DMD should also be tested.

From a general point of view, the problem faced in this work is exactly similar to the one in machine learning. We dispose of databases describing the flow we would like to predict (y) and the information that we have to make the prediction (x), and we are therefore looking for a model f that best predict y from x outside these databases. Many techniques have been developed in machine learning to find f under some hypothesis set chosen by the user. Radial Basis Function network is

one of them. Even if the results obtained in this work were not especially good, only one kind of RBF was tested and it is also possible to further specialize the shape parameters. It would be interesting to combine the physical knowledge of the flow (which will provide the hypothesis set) and machine learning techniques. In some sense, it is already done in this work, but only for what is probably the simplest machine learning method: the multivariate linear least square regression that we call Linear Stochastic Estimation.

As stated by Brunton et al. [102], machine learning has recently started to be used in turbulence and shown promising results. Lately, Nguyen [139] investigated the use of dictionary learning and Bayesian fusion for the estimation of small scales in turbulence. Bayesian fusion model were shown to be more accurate than LSE to enrich a database combining High-Time-Low-Space resolved and Low-Time-High-Space resolved data. However, in our case, only High-Time-Low-Space information is available and the method cannot be applied. The results obtained with dictionary learning were promising, but the method had to be careful handled to achieve those results. It is clear that pursuing the study of the applications of machine learning methods in the turbulence field is an important perspective of the present work.

Bibliography

[1] P. Jorna, R. Amalberti, N. McDonald, M. Piers, F. Taylor and A. Samel, "Safety in and around airports," Brussels, 1999.

[2] B. Herry, "Etude aérodynamique d'une double marche descendante 3D appliquée à la sécurisation de l'appontage des hélicoptères sur les frégates," Thèse de l'Université de Valenciennes, 2010.

[3] J. V. Healy, "The prospects for simulating the helicopter/ship interface," *Naval Engineers Journal*, vol. 99, no. 2, pp. 45-63, March 1987.

[4] R. J. Adrian, "On the role of conditional averages in turbulence theory," *Turbulence in liquids*, pp. 323-332, 1977.

[5] R. J. Adrian, "Conditional eddies in isotropic turbulence," *Physics of Fluids*, vol. 22, no. 11, pp. 2065-2070, November 1979.

[6] 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," *Physics of Fluids*, vol. 1, no. 6, pp. 992-998, June 1989.

[7] D. R. Cole, M. N. Glauser and Y. G. Guezennec, "An application of the stochastic estimation to the jet mixing layer," *Physics of Fluids*, vol. 4, no. 1, pp. 192-194, 1992.

[8] T. J. Giezeke and Y. G. Guezennec, "Stochastic Estimation of Multipoint Conditional Averages and their Spatio-Temporal Evolution," *Applied Scientific Research*, vol. 53, no. 3, pp. 305-320, December 1994.

[9] E. Vincendeau, "Analyse conditionnelle et estimation stochastique appliquées à l'étude des structures cohérentes dans la couche de mélange," Thèse de l'Université de Poitiers, 1995.

[10] W. C. Choi and Y. G. Guezennec, "On the asymmetry of structures in turbulent boundary layers," *Physics of Fluids*, vol. 2, no. 4, pp. 628-630, April 1990.

[11] N. E. Murray and L. S. Ukeiley, "Estimating the shear layer velocity field above an open cavity from surface pressure measurements," in *32nd AIAA Fluid Dynamics Conference and Exhibit*, Saint Louis, 2002, AIAA Paper 2002-2866.

[12] 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.

[13] 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.

[14] T. C. Tung and R. J. Adrian, "Higher-order estimates of conditional eddies is isotropic turbulence," *Physic of Fluids,* vol. 23, no. 7, pp. 1469-1470, July 1980.

[15] Y. G. Guezennec, "Stochastic estimation of coherent structures in turbulent boundary layers," *Physics of Fluids,* vol. 6, no. 1, pp. 1054-1060, 1989.

[16] A. Naguib, C. E. Warck and O. Juckenhofel, "Stochastic estimation and flow sources associated with surface pressure events in a turbulent boundary layer," *Physics of Fluids,* vol. 13, no. 9, pp. 2611-2626, 2001.

[17] J. P. Bonnet, D. R. Cole, J. Delville, M. N. Glauser and L. S. Ukeiley, "Stochastic estimation and proper orthogonal decomposition: complementary techniques for identifying structure," *Experiments in Fluids,* vol. 17, pp. 307-314, 1994.

[18] C. E. Tinney, F. Coiffet, J. Delville, A. M. Hall, P. Jordan and M. N. Glauser, "On spectral linear stochastic estimation," *Experiments in Fluids,* vol. 41, pp. 763-775, 2006.

[19] A. Sinha, A. Serrani and M. Samimy, "Development of Empirical Estimators for Feedback Control of High-Speed Axisymmetric Jets," *AIAA Journal,* vol. 49, no. 9, pp. 1971-1987, September 2011.

[20] V. Durgesh 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, no. 3, pp. 571-583, February 2010.

[21] C. W. Rowley, V. Juttijudata and D. R. Williams, "Cavity Flow Control Simulations and Experiments," in *43rd AIAA Aerospace Sciences Meeting,* Reno, 2005, AIAA Paper 2005-292.

[22] 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.

[23] 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.

[24] P. Mokhasi, D. Rempfer and S. Kandala, "Predictive flow-field estimation," *Physica D: Nonlinear Phenomena,* vol. 238, no. 3, pp. 290-308, February 2009.

[25] R. J. Adrian, "Stochastic estimation of organized turbulent structures: homogeneous shear flows," *Journal of Fluid Mechanics,* vol. 190, pp. 531-559, 1988.

[26] R. J. Adrian, "Stochastic Estimation of Conditional Structure: a Review," *Applied Scientific Research,* vol. 53, no. 3, pp. 291-303, December 1994.

[27] R. J. Adrian, Stochastic Estimation of the Structure of Turbulent Fields. In Eddy structure identification, J. P. Bonnet, Ed., Springer-Verlag, 1996.

[28] R. F. Dekou Tiomajou, "Organisation à grandes échelles de la turbulence de paroi," Thèse de l'Ecole Centrale Lille, Lille, 2016.

[29] G. J. Brereton, "Stochastic estimation as a statistical tool for approximating turbulent conditional averages," *Physics of Fluids,* vol. 4, no. 9, pp. 2046-2054, September 1992.

[30] M. G. Olsen and J. C. Dutton, "Stochastic Estimation of Large Structures in an Incompressible Mixing Layer," *AIAA Journal,* vol. 40, no. 2, pp. 2431-2438, December 2002.

[31] C. Picard and J. Delville, "Pressure velocity coupling in a subsonic round jet," *International Journal of Heat and Fluid Flow,* vol. 21, no. 3, pp. 359-364, June 2000.

[32] E. J. Gutmark, S. Verfaillie, J.-P. Bonnet and F. Grinstein, "Linear Stochastic Estimation of a Swirling Jet," *AIAA Journal,* vol. 44, no. 3, pp. 457-468, 2006.

[33] P. Druault, M. Yu and P. Sagaut, "Quadratic stochastic estimation of far-field acoustic pressure with coherent structure events in a 2D compressible plane mixing layer," *International Journal for Numerical Methods in Fluids,* vol. 62, pp. 906-926, 2010.

[34] F. Nicoud, J. S. Baggett, P. Moin and W. Cabot, "Large Eddy simulation wall-modeling based on suboptimal control theory and linear stochastic estimation," *Physics of Fluids,* vol. 13, no. 10, pp. 2968-2984, 2001.

[35] N. E. Murray and L. S. Ukeiley, "Estimation of the Flowfield from Surface Pressure Measurements in an Open Cavity," *AIAA Journal,* vol. 41, no. 5, pp. 969-972, 2003.

[36] N. Murray and L. Ukeiley, "Modified quadratic stochastic estimation of resonating subsonic cavity flow," *Journal of Turbulence,* vol. 8, no. 53, pp. 1-23, December 2007.

[37] N. E. Murray and L. S. Ukeiley, "Low-dimensional estimation of cavity flow dynamics," in *42nd AIAA Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 2004, AIAA Paper 2004-681.

[38] L. Ukeiley and N. Murray, "Velocity and surface pressure measurements in an open cavity," *Experiments in Fluids,* vol. 38, no. 5, pp. 656-671, 2005.

[39] E. Caraballo, J. Little, M. Debiasi and M. Samimy, "Developement and Implementation of an Experimental-Based Reduced-Order Model for Feedback Control of Subsonic Cavity Flows," *Journal of Fluids Engineering,* vol. 129, no. 7, pp. 813-824, July 2007.

[40] S. Arunajatesan, C. Kannepalli and L. S. Ukeiley, "Three Dimensional Stochastic Estimation Applied to Cavity Flow Fields," in *37th AIAA Fluid Dynamics Conference and Exhibit*, Miami, 2007, AIAA Paper 2007-4227.

[41] J. Kastner, D. Cuppoletti, E. Gutmark, A. Fahrland, J. B. Jeffries and R. K. Hanson, "Simultaneous Measurement of Flow Fluctuations and Near-Field Pressure in a Subsonic Jet," in *15th AIAA/CEAS Aeroacoustics Conference (30th AIAA Aeroacoustics Conference)*, Miami, Florida, 2009, AIAA Paper 2009-3211.

[42] J. L. Lumley, "The structure of inhomogeneous turbulent flows," *Atmosphere turbulence and radio propagation,* pp. 166-178, 1967.

[43] J. P. Bonnet, J. Delville, M. N. Glauser, R. A. Antonia, D. K. Bisset, D. R. Cole, H. E. Fiedler, J. H. Garem, D. Hilberg, J. Jeong, N. K. R. Kevlahan, L. S. Ukeiley and E.

Vincendeau, "Collaborative testing of eddy structure identification methods in free turbulent shear flows," *Experiments in Fluids,* vol. 25, no. 3, pp. 197-225, August 1998.

[44] J. H. Citriniti and W. K. George, "Reconstruction of the global velocity field in the axisymmetric mixing layer utilizing the proper orthogonal decomposition," *Journal of Fluid Mechanics,* vol. 418, no. 1, pp. 137-166, September 2000.

[45] P. B. V. Johansson and W. K. George, "The far downstream evolution of the high-Reynolds-number axisymmetric wake behind a disk. Part 2. Slice proper orthogonal decomposition," *Journal of Fluid Mechanics,* vol. 555, pp. 387-408, May 2006.

[46] A. S. Cruz, L. David, J. Pécheux and A. Texier, "Characterization by proper-orthogonal-decomposition of the passive controlled wake flow downstream of a half cylinder," *Experiments in Fluids,* vol. 39, no. 4, pp. 730-742, October 2005.

[47] G. Berkooz, J. Elezgaray, P. Holmes, J. Lumley and A. Poje, "The Proper Orthogonal Decomposition, Wavelets and Modal Approaches to the Dynamics of Coherent Structures," *Applied Scientific Research,* vol. 53, no. 3, pp. 321-328, December 1994.

[48] L. Perret, "Etude du couplage instationnaire calculs-expériences en écoulement turbulents," Thèse de l'Université de Poitiers, 2004.

[49] J. E. Spitler, J. W. Naughton and W. R. Lindberg, "An LSE/POD Estimation of the Wind Turbine Inflow Environnement Using Sparse Data," in *44th AIAA Aerospace Sciens Meeting and Exhibit,* Reno, 2006, AIAA Paper 2006-1364.

[50] C. E. Tinney, L. S. Ukeiley and M. N. Glauser, "Low-dimensional characteristics of a transonic jet. Part 2. Estimate and far-field prediction," *Journal of Fluid Mechanics,* vol. 615, pp. 53-92, November 2008.

[51] J. Borée, "Extended proper orthogonal decomposition: a tool to analyse correlated events in turbulent flows," *Experiments in Fluids,* vol. 35, pp. 188-192, 07 2003.

[52] C. Hoarau, J. Borée, J. Laumonier and Y. Gervais, "Analysis of the wall pressure trace downstream of a separated region using extended proper orthogonal decomposition," *Physics of Fluilds,* vol. 18, 2006.

[53] D. Ewing and J. H. Citriniti, "Examination of a LSE/POD complementary technique using single and multi-time information in the axisymmetric shear layer," in *IUTAM Symposium on Simulation and Identification of Organized Structures in Flows,* 1999.

[54] F. Coiffet, "Etude statistique du champ à proximité des jets axisymétriques turbulent à haut nombre de Reynolds," Thèse de l'Université de Poitiers, 2006.

[55] J. W. Hall and D. Ewing, "A Combined Spatial and Temporal Decomposition of the Coherent Structures in the Three-Dimensional Wall Jet," in *44th AIAA Aerospace Sciences Meeting and Exhibit,* Reno, Nevada, 2006, AIAA Paper 2006-308.

[56] J. W. Hall and D. Ewing, "Spectral Linear Stochastic Estimation of the Turbulent Velocity in a Square Three-Dimensional Wall Jet," *Journal of Fluids Engineering,* vol. 132, no. 5, May 2010.

[57] W. J. Baars and C. E. Tinney, "Proper orthogonal decomposition-based spectral higher-order stochastic estimation," *Physics of Fluids,* vol. 26, May 2014.

[58] L. Ukeiley, N. Murray, Q. Song and L. Cattafesta, "Dynamic Surface Pressure Based Estimation for Flow Control," in *In: IUTAM symposium on flow control and MEMS*, Berlin, 2008.

[59] T. Ruiz, C. Sicot, L. E. Brizzi, J. Borée and Y. Gervais, "Pressure/Velocity coupling induced by a near wall wake," *Experiments in Fluids,* vol. 49, no. 1, pp. 147-165, January 2010.

[60] M. Debiasi, J. Little, E. Carabello, X. Yuan, A. Serrani, J. H. Myatt and M. Samimy, "Influence of Model and Stochastic Estimation on the Control of Subsonic Cavity Flow – A Preliminary Study," in *3rd AIAA Flow Control Conference*, San Francisco, California, 2006, AIAA Paper 2006-3492.

[61] J. T. Pinier, J. M. Ausseur, M. N. Glauser and H. Higuchi, "Proportional Closed-Loop Feedback Control of Flow Separation," *AIAA Journal,* vol. 45, no. 1, pp. 181-190, January 2007.

[62] O. Stalnov, V. Palei, I. Fono, K. Cohen and A. Seifert, "Experimental estimation of a D-shaped cylinder wake using body-mounted sensors," *Experiments in Fluids,* vol. 42, pp. 531-542, 2007.

[63] K. Cohen, S. Siegel and T. McLaughlin, "A heuristic approach to effective sensor placement for modeling of a cylinder wake," *Computers and Fluids,* vol. 35, no. 1, pp. 103-120, 2006.

[64] D. Lasagna, M. Orazi and G. Iuso, "Multi-time delay, multi-point linear stochastic estimation of a cavity shear layer velocity from wall-pressure measurements," *Physics of Fluids,* vol. 25, 2013.

[65] 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.

[66] 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.

[67] D. Lasagna, L. Fronges, M. Orazi and G. Iuso, "Stochastic estimation of cavity shear layer and fully developed turbulent channel flows: linear and nonlinear multi-time-delay analyses," in *52nd Aerospace Sciences Meeting*, National Harbor, 2014, AIAA Paper 2014-0897.

[68] 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.

[69] P. Druault and P. Guibert, "Use of turbulent flow statistical properties for correcting erroneous velocity vectors in PIV," *Comptes Rendus de Mécanique,* vol. 332, no. 9, pp. 737-736, September 2004.

[70] L. H. Benedict and R. D. Gould, "Toward better uncertainty estimates for turbulence statistics," *Experiments in Fluids,* vol. 22, no. 2, pp. 129-136, 1996.

[71] L. Sirovich, "Turbulence and the dynamics of coherent structures, part1: Coherent structures," *Quarterly of Applied Mathematics,* vol. 45, no. 3, pp. 567-571, 1987.

[72] Y. S. Abu-Mostafa, M. Magdon-Ismail and H.-T. Lin, Learning From Data, AMLBook, 2012.

[73] A. K. Prasad and R. J. Adrian, "Stereoscopic particle image velocimetry applied to liquid flows," *Experiments in Fluids,* vol. 15, no. 1, pp. 49-60, 1993.

[74] G. E. Elsinga, F. Scarano, B. Wieneke and B. W. Van Oudheusden, "Tomographic particle image velocimetry," *Experiments in Fluids,* vol. 41, no. 6, pp. 933-947, 2006.

[75] C. Geiler and J. C. Monnier, "Industrialisation du Logiciel PIV - Phase 2," ONERA, Lille, 2012.

[76] E. W. Adams and J. P. Johnston, "Effects of the separating shear layer on the reattachment flow structure Part 1: Pressure and turbulence quantities," *Experiments in Fluids,* vol. 6, no. 6, pp. 400-408, 1988.

[77] P. G. Spazzini, G. Iuso, M. Onorato, N. Zurlo and G. M. Di Cicca, "Unsteady behavior of back-facing step flow," *Experiments in Fluids,* vol. 30, no. 5, pp. 551-561, 2001.

[78] J. Dandois, E. Garnier and P. Sagaut, "Numerical simulation of active separation control by a synthetic jet," *Journal of Fluid Mechanics,* vol. 574, pp. 25-58, 2007.

[79] I. Lee and H. J. Sung, "Multiple-arrayed pressure measurements for investigation of the unsteady flow structure of a reattaching shear layer," *Journal of Fluid Mechanics,* vol. 463, pp. 377-402, 2002.

[80] 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.

[81] J. M. Sousa, "Turbulent flow around a surface-mounted obstacle using 2D-3C DPIV," *Experiments in Fluids,* vol. 33, no. 6, pp. 854-862, 2002.

[82] R. Martinuzzi and C. Tropea, "The Flow Around Surface-Mounted, Prismatic Obstacles Placed in a Fully Developed Channel Flow," *Journal of Fluids Engineering,* vol. 115, no. 1, pp. 85-92, 1993.

[83] D. A. Lyn, S. Einav, W. Rodi and J.-H. Park, "A laser-Doppler velocimetry study of ensemble-averaged characteristics of the turbulent near wake of a square cylinder," *Journal of Fluid Mechanics,* vol. 304, pp. 285-319, 1995.

[84] S. Krajnovic and L. Davidson, "Large-Eddy Simulation of the Flow Around a Bluff Body," *AIAA Journal,* vol. 40, no. 5, pp. 927-936, 2002.

[85] G. Iaccarino, A. Ooi, P. A. Durbin and M. Behnia, "Reynolds averaged simulation of unsteady separated flow," *International Journal of Heat and Fluid Flow,* vol. 24, no. 2, pp. 147-156, 2003.

[86] E. R. Meinders, K. Hanjalic and R. J. Martinuzzi, "Experimental Study of the Local Convection Heat Transfer From a Wall-Mounted Cube in Turbulent Channel Flow," *Journal of Heat Transfer,* vol. 121, no. 3, pp. 564-573, 1999.

[87] A. Yakhot, H. Liu and N. Nikitin, "Turbulent flow around a wall-mounted cube: A direct numerical simulation," *Internation Journal of Heat and Fluid Flow,* vol. 27, no. 6, pp. 994-1009, 2006.

[88] L. Cambier, S. Heib and S. Plot, "The Onera elsA CFD Software: Input from Research and Feedback from Industry," *Mechanics and Industry,* vol. 14, no. 3, pp. 159-174, 2013.

[89] V. Brunet, "Computational Study of Buffet Phenomenon with Unsteady RANS Equations," in *21st AIAA Applied Aerodynamic Conference*, Orlando, Florida, 2003, AIAA Paper 2003-3679.

[90] V. Brunet and S. Deck, "Zonal-Detached Eddy Simulation of Transonic Buffet on a Civil Aircraft Type Configuration," in *38th AIAA Fluid Dynamics Conference and Exhibit*, Seattle, Washington, 2008.

[91] A. Garcia-Sagrado and T. Hynes, "Stochastic estimation of flow near the trailing edge of a NACA0012 airfoil," *Experiments in Fluids,* vol. 51, pp. 1057-1071, October 2011.

[92] E. Caraballo, T. Sullivan, R. You and J. Little, "Characterization of the Flow Field over a NACA 0015 Airfoil using Stochastic Estimation based on Surface Pressure and Hot Film Measurements," in *52nd Aerospace Sciences Meeting*, National Harbor, 2014, AIAA Paper 2014-0252.

[93] S. Deck, "Recent improvments in the Zonal Detached Eddy Simulation (ZDES) formulation," *Theoritical Computational Fluid Dynamics,* vol. 26, pp. 523-550, 2012.

[94] H. Toubin, "Prediction and Phenomenological Breakdown of Drag for Unsteady Flows," Thèse de l'Université Pierre et Marie Curie, Paris, 2015.

[95] K. B. Chun and H. J. Sung, "Control of turbulent separated flow over a backward-facing step by local forcing," *Experiments in Fluids,* vol. 21, no. 6, pp. 417-426, 1996.

[96] S. Pope, Turbulent flows, Cambridge University Pres, 2000.

[97] R. Perrin, M. Braza, E. Cid, S. Cazin, F. Thiele and J. Borée, "Time resolved stereoscopic PIV measurements in the near wake of a circular cylinder at high Reynolds number," in *14th International symposium on applications of laser techniques to fluid mechanics*, Lisbon, 2008.

[98] 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.

[99] K. Cohen, S. Siegel, M. Luchtenburg, T. McLaughlin and A. Seifert, "Sensor Placement for Closed-Loop Flow Control of a "D" Shaped Cylinder Wake," in *2nd AIAA Flow Control Conference*, Portland, Oregon, 2004, AIAA Paper 2004-2523.

[100] E. N. Lorenz, "Deterministic Nonperiodic Flow," *Journal of Atmospheric Sciences,* vol. 20, no. 2, pp. 130-141, 1963.

[101] N. Aubry, P. Holmes, J. L. Lumley and E. Stone, "The dynamics of coherent structures in the wall region of a turbulent boundary layer," *Journal of Fluid Mechanics,* vol. 192, pp. 115-173, 1988.

[102] S. L. Brunton and B. R. Noack, "Closed-Loop Turbulence Control: Progess and Challenges," *Applied Mechanics Review,* vol. 67, 2015.

[103] M. Balajewicz, F. Nitzsche and D. Feszty, "Application of Multi-Input Volterra Theory to Nonlinear Multi-Degree-of-Freedom Aerodynamic Systems," *AIAA Journal,* vol. 48, no. 1, pp. 56-62, 2010.

[104] D. J. Lucia, P. S. Beran and W. A. Silva, "Reduced-Order Modelling: New Approaches for Computational Physics," *Progress in Aerospace Sciences,* vol. 1, no. 2, pp. 51-117, 2004.

[105] B. R. Noack, P. Papas and P. A. Monkewitz, "The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows," *Journal of Fluid Mechanics,* vol. 523, pp. 339-365, 2005.

[106] B. R. Noack, I. Mezic, G. Tadmor and A. Banaszuk, "Optimal mixing in recirculation zones," *Physics of Fluids,* vol. 16, pp. 867-88, 2004.

[107] B. R. Noack, M. Morzynski and G. Tadmor, Reduced-Order Modelling for Flow Control, Springer-Verlag Wien, 2011.

[108] A. E. Deane, I. G. Kevrekidis, G. E. Karniadakis and S. A. Orszag, "Low-dimensional models for complex geometry flows: application to grooved channels and circular cylinders," *Physics of Fluids,* vol. 3, no. 2, pp. 2337-2354, 1991.

[109] L. Ukeiley, L. Cordier, R. Manceau, J. Delville, M. Glauser and J. P. Bonnet, "Examination of large-scale structures in a turbulent plane mixing layer. Part 2. Dynamical systems model," *Journal of Fluid Mechanics,* vol. 441, pp. 67-108, 2001.

[110] B. R. Noack and H. Eckelmann, "A low-dimensional Galerkin method for the three-dimensioncal flow around a circular cylinder," *Physics of Fluids,* vol. 6, no. 1, pp. 124-143, 1994.

[111] L. Mathelin and O. P. Le Maître, "Robust control of uncertain cylinder wake flows based on robust order models," *Computers and Fluids,* vol. 38, no. 6, pp. 1168-1182, 2009.

[112] D. Rempfer, "Investigation of boundary layer transition via Galerkin projection of empirical eigenfunctions," *Physics of Fluids,* vol. 8, no. 1, pp. 175-188, 1996.

[113] D. Rempfer, "On Low-Dimensional Galerkin Models for Fluid Flow," *Theoritical and Computational Fluid Dynamics,* vol. 14, pp. 75-88, 2000.

[114] B. R. Noack, K. Afanasiev, M. Morzynski, G. Tadmor and F. Thiele, "A hierarchy of low-dimensional models for the transient and post-transient cylinder wake," *Journal of Fluid Mechanics,* vol. 497, pp. 335-363, 2003.

[115] M. Couplet, P. Sagaut and C. Basdevant, "Intermodal energy transfers in a proper orthogonal decomposition-Galerkin representation of a turbulent separated flow," *Journal of Fluid Mechanics,* vol. 491, pp. 275-284, 2003.

[116] M. Couplet, C. Basdevant and P. Sagaut, "Calibrated reduced-order POD-Galerkin system for fluid flow modelling," *Journal of Computational Physics,* vol. 207, pp. 192-220, 2005.

[117] O. P. Le Maître and L. Mathelin, "Equation-free model reduction for complex dynamical systems," *International Journal for numerical methods in fluids,* vol. 63, pp. 163-184, 2010.

[118] L. Cordier, B. R. Noack, G. Tissot, G. Lehnasch, J. Delville, M. Balajewicz, G. Daviller and R. K. Niven, "Identification strategies for model-based control," *Experiments in Fluids,* vol. 54, no. 8, pp. 1-21, 2013.

[119] K. Wilcox and J. Peraire, "Balanced Model Reduction via the Proper Orthogonal Decomposition," *AIAA Journal,* vol. 40, no. 11, pp. 2323-2330, 2002.

[120] C. W. Rowley, "Model reduction for fluids using balanced proper orthogonal decomposition," *International Journal of Bifurcation and Chaos,* vol. 15, no. 3, pp. 997-1013, 2005.

[121] P. J. Schmid and D. S. Hennigson, Stability and Transition in Shear Flows, New York: Springer, 2001.

[122] G. Tadmor, O. Lehmann, B. R. Noack and M. Morzynski, "Mean field representation of the natural and actuated cylinder wake," *Physics of Fluids,* vol. 22, 2010.

[123] P. J. Schmid, "Dynamic Mode Decomposition of numerical and experimental data," *Journal of Fluid Mechnics,* vol. 656, pp. 5-28, 2010.

[124] C. W. Rowley, I. Mezic, S. Bagheri, P. Schlatter and D. Henningson, "Spectral analysis of nonlinear flows," *Journal of Fluid Mechnics,* vol. 641, pp. 115-129, 2009.

[125] E. Kaiser, B. R. Noack, L. Cordier, A. Spohn, M. Segond, M. Abel, G. Daviller, J. Osth, S. Krajnovic and R. K. Niven, "Cluster-based reduced-order modelling of a mixing layer," *Journal of Fluid Mechanics,* vol. 754, pp. 35-414, 2014.

[126] O. Talagrand, "Assimilation of Observations, an Introduction," *Journal of the Meteorological Society of Japan,* vol. 75, no. 1B, pp. 191-209, 1997.

[127] N. Papadakis, "Assimilation de données images : application au suivi de courbes et de champs de vecteurs," PhD thesis, Université Rennes I, 2007.

[128] J. Hoepffner, M. Chevalier, T. R. Bewley and D. S. Henningson, "State estimation in wall-bounded flow systems. Part 1. Perturbed laminar flows," *Journal of Fluid Mechanics,* vol. 534, pp. 263-294, 2005.

[129] M. Chevalier, J. Hoepffner, T. R. Bewley and D. S. Henningson, "State estimation in wall-bounded flow systems. Part 2. Turbulent flows," *Journal of Fluid Mechanics,* vol. 552, pp. 167-187, 2006.

[130] R. E. Kalman, "A new approach to linear filtering and prediction problems," *ASME-Journal of Basic Engineering,* vol. D, no. 82, pp. 35-45, 1960.

[131] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," University of North Carolina, Department of Computer Science, Chapel Hill, 2001.

[132] G. Evensen, Data Assimilation: The Ensemble Kalman Filter, Heidelberg: Springer-Verlag Berlin Heidelberg, 2009.

[133] G. Evensen, "Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical research,* vol. 99, no. C5, pp. 10143-10162, 1994.

[134] G. Evensen, "The Ensemble Kalman Filter: theoritical formulation and practical," *Ocean Dynamics,* vol. 53, no. 4, pp. 343-367, 2003.

[135] G. Burgers, P. J. van Leeuwen and G. Evensen, "Analysis scheme in the ensemble Kalma filter," *Monthly Weather Review,* vol. 126, pp. 1719-1724, 1998.

[136] L. Perret, E. Collin and J. Delville, "Polynomial identification of POD based low-order dynamical system," *Journal of Turbulence,* vol. 7, pp. 1-15, 2006.

[137] L. Cordier, "Flow control and constrained optimization problem," in *Reduced-Order Modelling for Flow Control,* B. R. Noack, M. Morzynski and G. Tadmor, Eds., Berlin, Springer, 2010, pp. 1-76.

[138] G. E. Fasshauer and J. G. Zhang, "On choosing "optimal" shape parameters for RBF approximation," *Numerical Algorithms,* vol. 45, no. 1, pp. 345-368, 2007.

[139] L. V. Nguyen, "Reconstruction of finely resolved velocity fields in turbulent flows from low resolution measurements," Université de Lille 1, Lille, 2016.

[140] W. J. Baars, C. E. Tinney and E. J. Powers, "POD based spectral Higher-Order Stochastic Estimation," in *48th AIAA Aerospace Sciences Meeting*, Orlando, 2010.

[141] J. Kastner, C. Harris and E. Gutmark, "Linear Stochastic Estimation of the Velocity Field from High-Bypass Nozzle with and without Mixing Devices," *Journal of Fluid Science and Technology,* vol. 6, no. 4, pp. 522-533, 2011.

[142] S. J. Wilkins, P. R. Richard and J. W. Hall, "Velocity Field Estimation Using Unsteady Wall Pressure Measurements in a Leading Edge Slat Flow," in *52nd Aerospace Sciences Meeting,* National Harbor, 2014, AIAA 2014-0896.

[143] A. Hekmati, D. Ricot and P. Druault, "About the convergence of POD and EPOD modes computed from CFD simulation," *Computers and Fluids,* vol. 50, pp. 60-71, 2011.

List of Tables

List of Figures

Annexes

A. Stochastic Estimation bibliography table

Study	Configuration	Experimental / Numerical	Method	Estimated data	Unconditional events	Goal
Adrian et al. (1979) [5]	Isotropic turbulence	Experimental	LSE and QSE	Velocity	Velocity	Structures identification / Flow dynamics study
Tung et al. (1980) [14]	Isotropic turbulence	Experimental	LSE and QSE	Velocity	Velocity	Structures identification / Flow dynamics study
Adrian et al. (1988) [25]	Homogeneous turbulent shear flow (DNS)	Numerical	LSE	Velocity	Velocity and deformation tensor	Structures identification / Flow dynamics study
Adrian et al. (1989) [6]	- Axisymmetric shear layer ($Re_d = 3.9\times10^5$ based on jet diameter) - Grid turbulence ($Re = 5.7\times10^4$ based on mesh spacing) - Plane shear layer (Velocity ratio: 0.53) - Pipe flow ($Re = 5\times10^4$)	Experimental	LSE and QSE	Velocity	Velocity	Structures identification / Flow dynamics study
Guezennec (1989) [15]	Fully turbulent boundary layer ($Re_\theta = 4.9\times10^3$ based on the momentum thickness at the measurements location)	Experimental	LSE and QSE	Velocity	Velocity	Structures identification / Flow dynamics study
Choi et al. (1989)	Flat-plate turbulent boundary layer ($Re_\theta = 4.9\times10^3$)	Experimental	LSE	Velocity	Velocity	Structures identification / Flow dynamics study
Brereton (1992) [29]	Flat-plate turbulent boundary layer ($Re_\theta = 3.1\times10^3$)	Experimental	LSE, QSE, Laurent Series SE	Velocity	Velocity	Structures identification Accuracy assessment in approximating conditional averages
Cole et al. (1992) [7]	Axisymmetric jet shear layer ($Re_d = 1\times10^5$)	Experimental	LSE	Velocity	Velocity	Structures identification / Flow dynamics study
Adrian (1994) [26]	Review of applications	Experimental				Structures identification / Flow dynamics study
Giezeke et al. (1994) [8]	Turbulent wake of a cylinder ($Re_d = 5\times10^3$)	Experimental	LSE	Velocity	Velocity	Structures identification / Flow dynamics study

Bonnet et al. (1994) [17]	- Axisymmetric jet shear layer ($Re_d = 1.1\times10^5$) - Plane shear layer (Velocity ratio: 0.64)	Experimental	Complementary technique	Velocity	Velocity	Structures identification / Flow dynamics study
Vincendeau (1995) [9]	Plane shear layer	Experimental	LSE	Velocity	Velocity	Structures identification / Flow dynamics study
Adrian (1996) [27]	Review of applications					Structures identification General description of SE
Bonnet et al. (1998) [43]	Plane shear layer (Velocity ratio: 0.6)	Experimental	Complementary technique	Velocity	Velocity	Structures identification / Flow dynamics study
Ewing et al. (1999) [53]	Axisymmetric jet shear layer ($Re_d = 8\times10^5$)	Experimental	Complementary technique in temporal and spectral domain	Velocity	Velocity	Low-dimensional reconstructions
Picard et al. (2000) [31]	Axisymmetric jet shear layer ($Re_d = 4.5\times10^4$)	Experimental	LSE POD	Velocity	Pressure	Structures identification / Flow dynamics study
Naguib et al. (2001) [16]	Flat-plate turbulent boundary layer ($Re_\theta = 1437 - 5670$)	Experimental	LSE, QSE	Velocity	Pressure	Study of the ability of LSE and QSE to approximate conditional average using surface pressure measurements
Nicoud et al. (2001) [34]	Turbulent channel flow ($Re_\tau = 640 - 2\times10^4$)	Numerical	LSE	Wall stress	Velocity	Generate wall stress model for LES simulations
Murray et al. (2002) [11]	Cavity flow (M = 1.5, length-to-depth ratio: 6)	Numerical	LSE, QSE	Velocity	Pressure	Flow reconstruction
Olsen et al. (2002) [30]	Plane shear layer (Velocity ratio: 0.575)	Experimental	LSE	Velocity	Deformation tensor	Structures identification / Flow dynamics study
Murray et al. (2003) [35]	Cavity flow (M = 1.5, length-to-depth ratio: 6)	Numerical	LSE, QSE	Velocity	Pressure	Flow reconstruction
Borée et al. (2003) [51]	-	-	EPOD			Theoretical demonstration of EPOD

Druault et al. (2004) [69]	Spark ignition engine flow	Experimental	LSE	Velocity	Velocity	Correction of PIV measurements
Perret (2004) [48]	Plane shear layer (Velocity ratio: 0.67)	Experimental	Complementary technique	Velocity	Velocity	Generation of inlet conditions for LES
Murray et al. (2004) [37]	Cavity flow (M = 0.2, length-to-depth ratio: 5.16 and 1.49)	Experimental	QSE, POD	Velocity	Pressure	Structures identification / Flow dynamics study
Taylor et al. (2004) [12]	Backward facing ramp with adjustable flap ($Re_h = 2\times10^4$)	Experimental	LSE-POD	Velocity	Pressure	Flow control
Rowley et al. (2005) [21]	2D cavity flow (M = 0.6, length-to-depth ratio: 2)	Numerical	LSE Dynamic observer	Velocity	Pressure	Flow control / Flow reconstruction
Ukeiley et al. (2005) [38]	Cavity flow (M = 0.2, length-to-depth ratio: 5.16 and 1.49)	Experimental	QSE	Velocity	Pressure	Structures identification / Flow dynamics study
Coiffet (2006) [54]	Subsonic and supersonic axisymmetric jet (M = 0.3 and 1.4)	Experimental	Spectral LSE, POD	Velocity	Pressure	Structures identification / Flow physics study
Hoarau et al. (2006) [52]	Forward facing ramp ($Re_h = 1\times10^5$)	Experimental	EPOD	Velocity	Pressure	Structures identification / Flow dynamics study
Spitler et al. (2006) [49]	Wind turbine inflow field	Numerical	LSE	Velocity	Velocity	Flow reconstruction
Debiasi et al. (2006) [60]	Cavity flow (M = 0.3, length-to-depth ratio: 4)	Experimental	(MTD)-LSE-POD, (MTD)-QSE-POD, POD-ROM	Velocity	Pressure	Flow control / Flow reconstruction
Gutmark et al. (2006) [32]	Swirling jet	Experimental	LSE	Velocity	Pressure	Structures identification / Flow dynamics study
Hall et al. (2006) [55]	3D wall jet ($Re_h = 8.96\times10^4$)	Experimental	LSE and Spectral LSE	Velocity	Pressure (with POD)	Structures identification / Flow dynamics study
Tinney et al. (2006) [18]	Axisymmetric jet shear layer ($Re_d = 4\times10^5$ - 8×10^5)	Experimental	Spectral LSE	Velocity	Pressure	Structures identification / Flow dynamics study
Murray et al. (2007) [36]	Cavity flow (M = 0.19, 0.29, 0.39, 0.58, 0.73, length-to-depth ratio: 6)	Experimental	QSE-POD	Velocity, density gradient	Pressure	Flow reconstruction

Hudy et al. (2007) [13]	Backward facing step (Re_h = 8081)	Experimental	LSE, QSE	Velocity	Pressure (with POD)	Structures identification / Flow dynamics study
Arunajatesan et al. (2007) [40]	Cavity flow (M = 2, length-to-depth ratio: 5)	Numerical	LSE-POD, QSE-POD	Velocity	Pressure	Structures identification / Flow dynamics study
Pinier et al. (2007) [61]	NACA-4412 ($Re_c = 1.35\times10^5$)	Experimental	LSE-POD, QSE-POD	Velocity	Pressure	Flow control / Flow reconstruction
Stalnov et al. (2007) [62]	D-shaped cylinder wake (Re_h = 225)	Experimental	LSE-POD	Velocity	Pressure	Flow control / Flow reconstruction
Ukeiley et al. (2008) [58]	Cavity flow (M = 0.6, length-to-depth ratio: 6)	Experimental	QSE-POD MTD-LSE Auto-Regressive Moving Average filter	Velocity	Pressure	Flow control / Flow reconstruction
Tinney et al. (2008) [50]	Axisymmetric jet (M = 0.85, d = 50.8mm)	Experimental	LSE-POD	Velocity	Pressure (with POD)	Structures identification / Flow dynamics study
Kastner et al. (2009) [41]	Axisymmetric nozzle jet and chevron nozzle jet	Experimental	LSE, QSE	Velocity	Pressure	Structures identification / Flow dynamics study
Ruiz et al. (2010) [59]	Wake of a disk located near a flat wall ($Re_d = 1.3\times10^5$ based on disk diameter)	Experimental	LSE and ST-POD	Velocity	Pressure	Structures identification / Flow dynamics study
Nguyen et al. (2010) [22]	Backward facing step (Re_h = 8081)	Experimental	LSE-POD, PCR-POD, KRR-POD and their MTD version	Velocity	Wall shear	Flow reconstruction
Baars et al. (2010) [140]	Axisymmetric jet	Numerical	POD based Spectral HOSE	Pressure	Pressure	Structures identification
Hall et al. (2010) [56]	3D wall jet ($Re_h = 8.96\times10^4$ based on jet channel)	Experimental	Spectral LSE	Velocity	Pressure	Structures identification / Flow dynamics study
Durgesh et al. (2010) [20]	Wake downstream a bluff body ($Re_h = 2\times10^4$)	Experimental	MTD-LSE-POD	Velocity	Pressure	Flow reconstruction
Druault et al. (2010) [33]	Plane shear layer (Velocity ratio: 0.5)	Numerical	QSE and POD	Pressure	Velocity	Structures identification / Flow dynamics study

Sinha et al. (2011) [19]	Axisymmetric jet ($Re_d = 3.6\times10^4$)	Numerical	LSE-POD, QSE-POD, KF, EKF, LTIF	Velocity	Pressure	Flow control / Flow reconstruction
Garcia-Sagrado et al. (2011) [91]	Wake downstream a NACA0012 profile ($Re_c = 2\times10^5$)	Experimental	LSE and spectral LSE	Velocity	Pressure	Structures identification / Flow dynamics study
Kastner et al. (2011) [141]	Coaxial nozzle (Center flow: M = 0.96, Stagnation temperature = 377°K; Coaxial flow: M = 0.5, Stagnation temperature = 300°K)	Experimental	LSE	Velocity	Velocity	Structures identification / Flow dynamics study
Tu et al. (2012) [65]	Wake downstream a bluff body ($Re_h = 3.6\times10^3$)	Experimental	MTD-LSE-POD, Kalman smoother, KF	Velocity	Velocity	Flow reconstruction
Lasagna et al. (2013) [64]	Cavity flow (with curvilinear geometry) (length-to-depth ratio: 6.6)	Experimental	LSE, QSE, MTD-LSE	Velocity	Pressure	Flow reconstruction
Carabello et al. (2014) [92]	Flow field over a NACA0015 ($Re_c = 1.15\times10^6$)	Experimental	MTD-LSE-POD	Velocity	Pressure/Velocity	Structures identification / Flow dynamics study
Lasagna et al. (2014) [67]	- Cavity flow (with curvilinear geometry) (U = 5.8 $m.s^{-1}$, length-to-depth ratio: 6.6) - Turbulent channel flow ($Re_\tau = 180$)	Experimental	MTD-LSE Nonlinear MTD-SE	Velocity	- Pressure - Shear stress	Flow reconstruction
Wilkins et al. (2014) [142]	Leading edge slat flow ($Re_c = 6\times10^5$)	Experimental	LSE, LSE-POD	Velocity	Pressure	Structures identification / Flow dynamics study
Clark et al. (2014) [23]	Wake downstream a bluff body ($Re_h = 1.6\times10^4$)	Experimental	MTD-LSE-POD	Velocity	Pressure	Flow reconstruction
Baars et al. (2014) [57]	Axisymmetric jet	Numerical	POD based Spectral HOSE	Pressure	Pressure	Structures identification
Hosseini et al. (2015) [66]	Wake downstream a pyramid	Experimental	MTD-EPOD with antisym-	Velocity	Pressure	Flow reconstruction

			metric, symmetric flow decomposition			
Lasagna et al. (2015) [68]	- Cavity flow (with curvilinear geometry) (U = 5.8 m.s^{-1}, length-to-depth ratio: 6.6) - Turbulent channel flow (Re_τ = 180)	Experimental	MTD-LSE Non linear MTD-SE	Velocity	- Pressure - Shear stress	Flow reconstruction

Table A.1: Summary of Stochastic Estimation studies

B. Tikhonov regularization

As explained in Chapter II, finding the coefficients of the SE (linear, quadratic or multi-time-delayed) is equivalent to solving an overdetermined system of the form:

$$\mathrm{EA} = \mathrm{U} \tag{VI.20}$$

where E is the matrix of conditional events (or combination of conditional events for Higher-Order SE), A is the matrix of coefficients of the method and U is the matrix of data to estimate. This system is solved using Ordinary Least Square regression and the solution is: $\mathrm{A} = (\mathrm{E}\mathrm{E}^{\mathrm{T}})^{-1}\mathrm{E}^{\mathrm{T}}\mathrm{U}$. This solution is the one that minimizes the loss function *f* defined by the L_2-norm of the residual $\mathrm{EA} - \mathrm{U}$:

$$f(\mathrm{A},\mathrm{U}) = \|\mathrm{EA} - \mathrm{U}\|^2 \tag{B.1}$$

However, this system is ill-posed and the matrix E can be very ill-conditioned. In such situation, it may be better to compute an approximate solution using a nearby system that is less sensitive to perturbations than the initial one. To do so, the loss function is penalized. Using the Tikhonov regularization [72], the loss function is penalized using the L_2-norm and becomes:

$$f(A,U) = \|\mathrm{EA} - \mathrm{U}\|^2 + \|\mathrm{LA}\|^2 \tag{B.2}$$

Then the coefficients $\mathrm{A}(\boldsymbol{x})$ are solution of the following regularized system:

$$(\mathrm{M}^{\mathrm{T}}\mathrm{M} + \mathrm{L}^{\mathrm{T}}\mathrm{L})\mathrm{A}(\boldsymbol{x}) = \mathrm{M}^{\mathrm{T}}\mathrm{U}(\boldsymbol{x}) \tag{B.3}$$

where L is called the Tikhonov matrix. Most often, L is taken such as being proportional to the identity matrix: $\mathrm{L} = \alpha\mathrm{I}$. Such matrix is used in this work. And the regularization parameter α is determined by cross-validation.

In comparison with the OLS where A satisfies the minimization problem $\min_{\mathrm{A}(\boldsymbol{x})\in\mathbf{R}^{N.N_d}}(\|\mathrm{EA}(\boldsymbol{x}) - \mathrm{U}(\boldsymbol{x})\|^2)$, using the Tikhonov regularization A satisfies the minimization problem: $\min_{\mathrm{A}(\boldsymbol{x})\in\mathbf{R}^{N.N_d}}(\|\mathrm{EA} - \mathrm{U}\|^2 + \propto^2 \|\mathrm{U}(\boldsymbol{x})\|^2)$.

In our case, only one regularization parameter is used for the complete system formed with A the matrix containing all the columns $\mathrm{A}(\boldsymbol{x})$ for all $\boldsymbol{x}$ of the field to reconstruct and U all the columns $\mathrm{U}(\boldsymbol{x})$. When SE-POD is used, only one regularization parameter is also used for all the POD coefficients estimated.

C. About POD basis convergence

The convergence of the POD basis from the low speed PIV database of the flow downstream of the backward facing step is investigated using a methodology inspired by Hekmati et al. work [143]. First, the influence of the spatial discretization is assessed by comparing the POD modes obtained with the original PIV snapshots with the POD modes obtained with spatially downsampled PIV snapshots. The snapshots of the training set are spatially deteriorated so that only one every two mesh points in both streamwise and vertical direction is kept. Then, a new POD basis is computed using those PIV snapshots (and referred to as the "modified" basis). To compare the POD modes of two bases, the following correlation coefficient is used:

$$C_i = \frac{\left(\boldsymbol{\Phi}_{i,reference}(\boldsymbol{x}), \boldsymbol{\Phi}_{i,modified}(\boldsymbol{x})\right)_\Omega}{\sqrt{\left(\boldsymbol{\Phi}_{i,reference}(\boldsymbol{x}), \boldsymbol{\Phi}_{i,reference}(\boldsymbol{x})\right)_\Omega \left(\boldsymbol{\Phi}_{i,modified}(\boldsymbol{x}), \boldsymbol{\Phi}_{i,modified}(\boldsymbol{x})\right)_\Omega}} \quad \text{(C.1)}$$

where i is the rank of the POD mode considered and $(.,.)_\Omega$ the dot product on the spatial domain Ω. The correlation coefficients of the first 400 POD modes are plotted in Fig. C.1. Up to POD mode 70, correlation coefficients higher than 0.8 are obtained. For some of the 70 first POD modes, the correlation coefficient drops abruptly. When such drop occurs, it actually concerns two consecutive POD modes. Those drops are in fact due to a misclassification of the POD mode rank. For instance, POD modes 32 and 33 have correlation coefficients lower than 0.4. But, if the correlation coefficient is computed between the POD mode 32 of the reference basis and the POD mode 33 of the modified basis, then the correlation coefficient goes up to 0.92. At last, the TKE content of these two modes are 0.442% and 0.441% for both bases which is very close. In our opinion, these inversions are not the symptom of a bad convergence of these modes POD which spatial content is very similar in both bases. In overall, from Fig. C.1, it appears that the spatial discretization is good enough to consider the first 70 POD modes as converged.

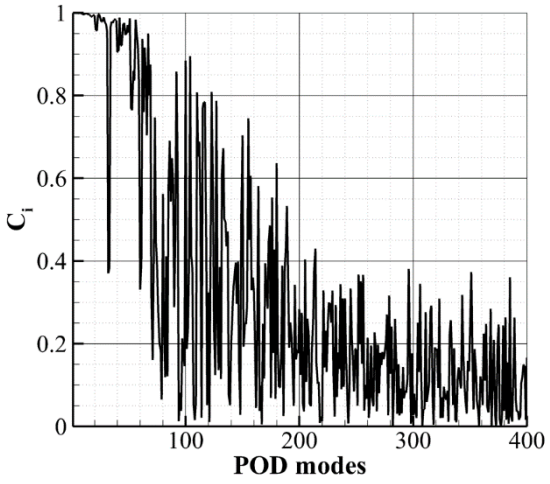


Fig. C.1: Correlation coefficients between POD modes obtained using the original PIV snapshots and the ones obtained using spatially downsampled PIV snapshots

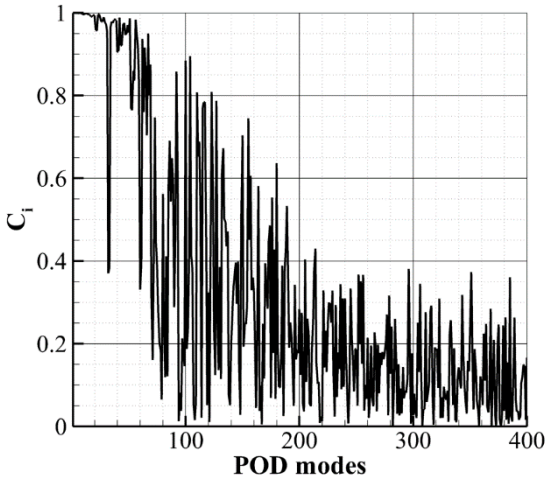


Fig. C.2: Correlation coefficients between POD modes obtained using the all original PIV snapshots and the ones obtained using only the training set PIV snapshots

A similar study is performed to assess the impact of the number of snapshots used to compute the POD basis. The reference basis is this time the one formed using all the PIV snapshots available (10 200 snapshots). This basis is compared with the POD basis obtained using only the training set data (6 794 snapshots). Results are shown in Fig. C.2. Once more, up to POD mode 80, it is possible to re-associate the POD modes for which C_i has dropped and obtain correlation coefficient higher than 0.85. The number of PIV snapshots contained in the training set is enough to consider at least the first 80 POD modes as converged. From these convergence studies, we are confident that the first 70 POD modes obtained from the training data set hold meaningful information on the flow physics and can be considered as converged.

Fig. C.3 shows the energy spectrum of the POD decomposition for the first six hundred modes, containing approximately 89% of the TKE. In addition, energy spectra of the two modified POD bases used to study the convergence of the POD are also plotted. The three spectra match well for these POD modes (it is also true up to the 2000th POD modes, at this point the spectrum of the

POD basis obtained from the downsampled PIV snapshots diverge from the other two spectra). Therefore, even if the spatial content of POD modes of rank higher than 70 is not converged, their energy content is converged. A POD basis better converged will have the same spectrum. The first 20 POD modes account for 50% of the TKE but, to retain 90% of the TKE, the first 650 POD modes have to be considered, and that is not the results of insufficient convergence of the basis. In these conditions, the flow does not appear to be dominated by highly energetic, large and coherent structures that can be described in a small number of POD modes and easily correlated with a limited number of conditional events.

Also, the inaccuracy of the prediction of high order POD modes (higher than rank 70), and thus their low correlations to the sensors, do not seem to be particularly related to the fact that they are not converged. The prediction of the POD modes of a POD basis calculated using only 3397 PIV snapshots was compared with the prediction of the POD modes of the training data set POD basis. When using only 3397 PIV snapshots, only the first 30 POD modes can be considered as converged. Nevertheless, similar determination coefficients are obtained between the two predictions even for POD modes of rank much higher than 30. In particular, the trends are not modified. To illustrate this phenomenon, R^2POD Single of the two predictions using the grid of 462 velocity sensors are shown in Fig. C.4. Up to the 20th POD modes, values obtained in both situations almost perfectly match each other. For POD modes of higher ranks, some discrepancies can be observed but the general trends and order of magnitude remain the same.

Therefore, from this study, it seems that results and conclusions given in this work are not distorted even if the spatial content of high order POD modes are not converged. The main concern is for the optimization of the sensor locations. When optimizing the sensor locations for one or a set of POD modes, the sensor locations are related to the spatial content of the POD modes and thus one should not use unconverged POD modes for the optimization.

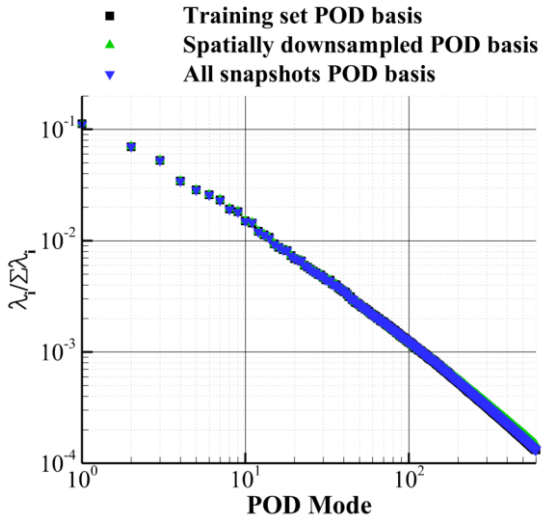


Fig. C.3: Energy spectrum of the first six hundred POD modes

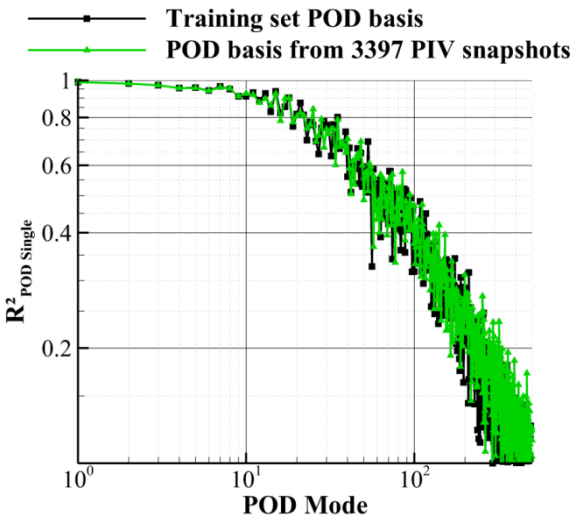


Fig. C.4: Determination coefficient of the estimation of POD modes from the training set POD basis and a POD basis formed using the first 3397 PIV snapshots of the training set (determination coefficients are evaluated on the validation set)

D. Turbulence integral length scales calculation

The turbulence spatial integral length scales of one velocity component are calculated from the spatial autocorrelation function according to equation [80]:

$$\Lambda(\boldsymbol{x_0}) = \int_0^{\infty} R_{ii}(\boldsymbol{x_0}, r)dr \tag{D.1}$$

where $\boldsymbol{x_0}$ is the position at which the length scale is calculated and the double-*i* subscript indicates the autocorrelation function of the component *i* defined by:

$$R_{ii}(\boldsymbol{x_0}, r) = \frac{\langle u_i(\boldsymbol{x_0}, t)u_i(\boldsymbol{x_0} + r, t)\rangle}{\sqrt{\langle u_i^2(\boldsymbol{x_0})\rangle\langle u_i^2(\boldsymbol{x_0} + r)\rangle}} \tag{D.2}$$

and *r* is the distance between two points in the flow. Following O'Neill et al., the integration domain of equation (D.1) was chosen to range from 0 to the point where the autocorrelation goes and remains under 1/*e*, instead of ranging from 0 to +∞.

In practice the turbulence spatial integral length scale is calculated on a Cartesian mesh, thus equation (D.1) becomes:

At last, we denote the integral length scale calculated in the streamwise direction, the streamwise integral length scale (whatever the velocity component considered). The vertical integral length scale denotes the integral length scale calculated in the vertical direction. Since the integral length scales are computed on a finite domain, with a criterion for stopping the integration, they may not be perfectly converged in some points of the field. Indeed, the spatial domain may not always be large enough for the autocorrelation function to reach and remain close to 0 (in particular near the edges of the domain but the combination of Λ^+ and Λ^- should limit this problem). Also, according to Pope [96], the spatial domain should be at least six times larger than the maximum integral length value to accurately determine such integral length scale. However, we are interested in the comparison of original values with estimated ones, not in the integral length scale value in itself. Convergence issues are not expected to false the comparison made in Chapter IV.3.

E. Link between Spatio-Temporal Linear Stochastic Estimation and Multi-Time-Delay Linear Stochastic Estimation

As described by Ruiz et al. [59], the Spatio-Temporal POD consists in computing the POD of a spatio-temporal set of data. Let us consider N time samples of some data of dimension M. The set of data is denoted $\mathrm{U} \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}} = \left(u_i(t_j)\right)$. Instead of computing the POD of the matrix U, the spatio-temporal POD computes the POD of an extended matrix $\mathrm{U_{ST}}$ which contains delayed data. Assuming that 10 delayed data in the past and future are used (with a step between delays of one sample) then the extended matrix is:

$$\mathrm{U_{ST}} = \begin{pmatrix} u_1(t_1) & \cdots & u_1(t_{\mathrm{N}-20}) \\ u_1(t_2) & \ddots & u_1(t_{\mathrm{N}-19}) \\ \vdots & \ddots & \vdots \\ u_1(t_{21}) & \ddots & u_1(t_{\mathrm{N}}) \\ u_2(t_1) & \ddots & u_2(t_{\mathrm{N}-20}) \\ \vdots & \ddots & \vdots \\ u_{\mathrm{M}}(t_{21}) & \cdots & u_{\mathrm{M}}(t_{\mathrm{N}}) \end{pmatrix} \in \mathbb{R}^{(21\times\mathrm{M})\times(\mathrm{N}-20)}$$

The ST-POD of U is the POD of $\mathrm{U_{ST}}$.

The Spatio-Temporal LSE proposed by Ruiz et al. is then to compute the LSE using as conditional events the ST-POD coefficients of the sensors. Thus, first a ST-POD of the sensors is computed (a number of delays as to be chosen and it is possible to consider delays only in the past), then LSE is perform using the ST-POD coefficients as conditional events (and it is possible to select the number of ST-POD coefficients to use). As Borée demonstrated that Extended POD is equivalent to LSE if all the extended POD modes are considered, we now demonstrate that the ST-LSE is equivalent to the MTD-LSE if all ST-POD coefficients are used as conditional events. The main advantage of the ST-LSE is then to dispose of a POD decomposition of the sensors enriched of delays. It is then possible to select a limited number of POD coefficients which can reduced the overfitting and represents an alternative to the use of Tikhonov Regularization in MTD-LSE.

Let us considered the matrix of conditional events composed of the sensors and their delays: $E \in \mathbb{R}^{(N_d \times N_S) \times N}$, with N_d the number of delays (synchroneous time included), N_S the number of sensors and N the number of time samples. The data to estimate are organized in the matrix $U \in \mathbb{R}^{N \times M}$, where M is the dimension of the data to estimate. The matrix of MTD-LSE coefficients is then $A_{MTD-LSE} = (EE^T)^{-1}EU \in \mathbb{R}^{(N_d \times N_S) \times M}$ and the estimation by MTD-LSE is $\widetilde{U}_{MTD-LSE} = E^T A_{MTD-LSE}$.

Now the ST-LSE uses the ST-POD coefficients of E as conditional events. Using POD, E can be decomposed in:

$$E = \boldsymbol{\Phi}_{ST-POD} M_{ST-POD} \tag{E.1}$$

with $\boldsymbol{\Phi}_{ST-POD}$ the POD modes of the decomposition of E $\left(\boldsymbol{\Phi}_{ST-POD} \in \mathbb{R}^{(N_d \times N_S) \times N_{ST-POD}}\right)$, M_{ST-POD} the POD coefficients ($M_{ST-POD} \in \mathbb{R}^{N_{ST-POD} \times N}$). $N_{ST-POD} = \min(N_d \times N_S, N)$.

Thus the ST-LSE coefficients are defined by:

$$A_{ST-LSE} = \left(M_{ST-POD} M_{ST-POD}^T\right)^{-1} M_{ST-POD} U \in \mathbb{R}^{(N_d \times N_S) \times M} \tag{E.2}$$

And the estimation by ST-LSE is:

$$\widetilde{U}_{ST-LSE} = M_{ST-POD}^T A_{ST-LSE} \tag{E.3}$$

We now want to demonstrate that $\widetilde{U}_{MTD-LSE} = \widetilde{U}_{ST-LSE}$.

Using (E.1) and the fact that $\boldsymbol{\Phi}_{ST-POD}^T P \boldsymbol{\Phi}_{ST-POD} = I$,

$$M_{ST-POD} = \boldsymbol{\Phi}_{ST-POD}^T PE \tag{E.4}$$

P is the dot product matrix used for computing the POD of E and is symmetric positive-definite.

By injecting (E.5) in (E.2),

$$A_{ST-LSE} = \left(\boldsymbol{\Phi}_{ST-POD}^T PEE^T P \boldsymbol{\Phi}_{ST-POD}\right)^{-1} \boldsymbol{\Phi}_{ST-POD}^T PEU$$

$$\boldsymbol{\Phi}_{ST-POD}^T PEE^T P \boldsymbol{\Phi}_{ST-POD} A_{ST-LSE} = \boldsymbol{\Phi}_{ST-POD}^T PEU$$

$$EE^T P \boldsymbol{\Phi}_{ST-POD} A_{ST-LSE} = EU$$

$$P \boldsymbol{\Phi}_{ST-POD} A_{ST-LSE} = A_{MTD-LSE} \tag{E.5}$$

Therefore, using (E.1), (E.5) and (E.3) respectively,

$$\widetilde{U}_{MTD-LSE} = E^T A_{MTD-LSE} = M_{ST-POD}^T \boldsymbol{\Phi}_{ST-POD}^T A_{MTD-LSE} = M_{ST-POD}^T \boldsymbol{\Phi}_{ST-POD}^T P \boldsymbol{\Phi}_{ST-POD} A_{ST-LSE}$$

$$\widetilde{\mathrm{U}}_{\mathrm{MTD-LSE}} = \mathrm{M}_{\mathrm{ST-POD}}^{\mathrm{T}}\mathrm{A}_{\mathrm{ST-LSE}} = \widetilde{\mathrm{U}}_{\mathrm{ST-LSE}}$$

In the demonstration, only training data were considered. The equivalence between the ST-LSE and MTD-LSE when all POD coefficients of the ST-POD are used is also true outside the training set. This can be demonstrated using $\mathrm{M}'_{\mathrm{ST-POD}} = \boldsymbol{\Phi}_{\mathrm{ST-POD}}^{\mathrm{T}}\mathrm{PE}'$ where E' are considitonal events outside the training set. The equation (C.1) still holds outside the training set and from these relation one easily gets that:

$$\widetilde{\mathrm{U}}_{\mathrm{MTD-LSE}} = \mathrm{E}'^{\mathrm{T}}\mathrm{A}_{\mathrm{MTD-LSE}} = \mathrm{M}'^{\mathrm{T}}_{\mathrm{ST-POD}}\mathrm{A}_{\mathrm{ST-LSE}} = \widetilde{\mathrm{U}}_{\mathrm{ST-LSE}}$$

F. Precisions on Kalman filter

1. General expression of Kalman filter gain

Let us consider a linear dynamic model $\boldsymbol{x}_{k+1}^{\mathrm{t}} = \mathrm{M}\boldsymbol{x}_k^{\mathrm{t}} + \boldsymbol{\eta}_k$ and a linear observation model $\boldsymbol{y}_k = \mathrm{H}\boldsymbol{x}_k^t + \boldsymbol{\varepsilon}_k$. $\boldsymbol{x} \in \mathbb{R}^n$ is the system state. $\boldsymbol{y} \in \mathbb{R}^p$ are some observations. M is the matrix of the linear dynamic model linking the state at time $t_{k+1} = t_k + \Delta t$ with the state at time t_k. H is the matrix of the observation model linking the state to the observations. To simplify the demonstration (but those assumptions are not restrictive), Δt, M and H are assumed to be constant over time. $\boldsymbol{\eta}_k$ and $\boldsymbol{\varepsilon}$ are the dynamic and observation model error respectively. They are supposed to be random variable following Gaussian distribution with no bias and covariance matrix Q and R respectively. Superscript "t" refers to the true state, subscripts refer to the time step.

As described in Chapter VI.2, the Kalman filter gain is expressed as (with $\mathrm{P}_{k+1}^{\mathrm{b}}$ the error covariance matrix of the Kalman filter predicted state, superscript "b" refers to the predicted state):

$$\mathrm{K}_{k+1} = \mathrm{P}_{k+1}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}}\left(\mathrm{H}\mathrm{P}_{k+1}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}} + \mathrm{R}\right)^{-1}$$

Let us now consider:

$$\mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}} = \mathrm{E}\left[\left(\boldsymbol{x}_{k+1}^{\mathrm{b}} - \boldsymbol{x}_{k+1}^{\mathrm{t}}\right)\left(\mathrm{H}\boldsymbol{x}_{k+1}^{\mathrm{b}} - \boldsymbol{y}_{k+1}^{t}\right)^{\mathrm{T}}\right]$$

$$\mathrm{P}_{\boldsymbol{y}\boldsymbol{y}_{k+1}}^{\mathrm{b}} = \mathrm{E}\left[\left(\mathrm{H}\boldsymbol{x}_{k+1}^{\mathrm{b}} - \boldsymbol{y}_{k+1}^{t}\right)\left(\mathrm{H}\boldsymbol{x}_{k+1}^{\mathrm{b}} - \boldsymbol{y}_{k+1}^{t}\right)^{\mathrm{T}}\right]$$

Then:

$$\begin{aligned}\mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}} &= \mathrm{E}\left[\boldsymbol{e}_{k+1}^{b}\left(\mathrm{H}\boldsymbol{x}_{k+1}^{\mathrm{b}} - \mathrm{H}\boldsymbol{x}_{k+1}^{\mathrm{t}} - \boldsymbol{\varepsilon}_{k+1}\right)^{\mathrm{T}}\right] = \mathrm{E}\left[\boldsymbol{e}_{k+1}^{b}\left(\boldsymbol{x}_{k+1}^{\mathrm{b}} - \boldsymbol{x}_{k+1}^{\mathrm{t}}\right)^{\mathrm{T}}\mathrm{H}^{\mathrm{T}}\right] - \mathrm{E}[\boldsymbol{e}_{k+1}^{b}(\boldsymbol{\varepsilon}_{k+1})^{\mathrm{T}}] \\ &= \mathrm{E}\left[\boldsymbol{e}_{k+1}^{b}\left(\boldsymbol{e}_{k+1}^{b}\right)^{\mathrm{T}}\right]\mathrm{H}^{\mathrm{T}} = \mathrm{P}_{k+1}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}}\end{aligned}$$

And:

$$\begin{aligned}\mathrm{P}_{\boldsymbol{y}\boldsymbol{y}_{k+1}}^{\mathrm{b}} &= \mathrm{E}\left[\left(\mathrm{H}\boldsymbol{x}_{k+1}^{\mathrm{b}} - \mathrm{H}\boldsymbol{x}_{k+1}^{\mathrm{t}} - \boldsymbol{\varepsilon}_{k+1}\right)\left(\mathrm{H}\boldsymbol{x}_{k+1}^{\mathrm{b}} - \mathrm{H}\boldsymbol{x}_{k+1}^{\mathrm{t}} - \boldsymbol{\varepsilon}_{k+1}\right)^{\mathrm{T}}\right] \\ &= \mathrm{H}\,\mathrm{E}\left[\boldsymbol{e}_{k+1}^{b}\left(\boldsymbol{e}_{k+1}^{b}\right)^{\mathrm{T}}\right]\mathrm{H}^{\mathrm{T}} + \mathrm{E}[\boldsymbol{\varepsilon}_{k+1}(\boldsymbol{\varepsilon}_{k+1})^{\mathrm{T}}] = \mathrm{H}\mathrm{P}_{k+1}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}} + \mathrm{R}\end{aligned}$$

Therefore, the Kalman gain can be expressed as:

$$\mathrm{K}_{k+1} = \mathrm{P}_{k+1}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}}\left(\mathrm{H}\mathrm{P}_{k+1}^{\mathrm{b}}\mathrm{H}^{\mathrm{T}} + \mathrm{R}\right)^{-1} = \mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}}\left(\mathrm{P}_{\boldsymbol{y}\boldsymbol{y}_{k+1}}^{\mathrm{b}}\right)^{-1}$$

In the Ensemble Kalman filter, Monte Carlo method is used to approximate the covariance error matrices and thus:

$$\mathrm{P}_{k+1}^{\mathrm{b}} \approx \frac{1}{\mathrm{N}-1}\sum_{i=1}^{\mathrm{N}}\left[\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}-\overline{\boldsymbol{x}_{k+1}^{\mathrm{b}}}\right)\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}-\overline{\boldsymbol{x}_{k+1}^{\mathrm{b}}}\right)^{\mathrm{T}}\right]$$

$$\mathrm{R}_{k+1}^{\mathrm{N}} = \frac{1}{\mathrm{N}-1}\sum_{i=1}^{\mathrm{N}}\left(\boldsymbol{\varepsilon}_{k+1}^{i}\right)\left(\boldsymbol{\varepsilon}_{k+1}^{i}\right)^{\mathrm{T}}$$

$$\mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}} \approx \frac{1}{\mathrm{N}-1}\sum_{i=1}^{N}\left[\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}-\overline{\boldsymbol{x}_{k+1}^{\mathrm{b}}}\right)\left(H\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}\right)-\overline{H\left(\boldsymbol{x}_{k+1}^{\mathrm{b}}\right)}\right)^{\mathrm{T}}\right]$$

$$\mathrm{P}_{\boldsymbol{y}\boldsymbol{y}_{k+1}}^{\mathrm{b}} \approx \frac{1}{\mathrm{N}-1}\sum_{i=1}^{N}\left[\left(H\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}\right)-\overline{H\left(\boldsymbol{x}_{k+1}^{\mathrm{b}}\right)}\right)\left(H\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}\right)-\overline{H\left(\boldsymbol{x}_{k+1}^{\mathrm{b}}\right)}\right)^{\mathrm{T}}\right] + \mathrm{R}_{k+1}^{\mathrm{N}}$$

where superscript "*i*" denotes the indice of the ensemble members in the Monte Carlo method. $\overline{\cdot}$ refers to the ensemble average.

2. "Inverse" observation model

"Inverse" observation model consists in considering $\boldsymbol{x}_k = H'(\boldsymbol{y}_k) + \boldsymbol{\varepsilon}'$ with $\boldsymbol{\varepsilon}'$ the "inverse" observation model error which is supposed to be a random variable following a Gaussian distribution with the covariance R′. In the Kalman filter equations, the equivalent observation model is formed with H being the identity and an observation error covariance matrix R′. The prediction steps remain unchanged. The analysis steps are modified as follow:

- Compute the Kalman gain: $\mathrm{K}_{k+1} = \mathrm{P}_{k+1}^{\mathrm{b}}\left(\mathrm{P}_{k+1}^{\mathrm{b}} + \mathrm{R}'\right)^{-1}$
- Compute the approximate observed state with the "inverse" model: $\boldsymbol{x}_k^{\mathrm{obs}} = H'(\boldsymbol{y}_k)$
- Generate the analyzed state: $\boldsymbol{x}_{k+1}^{\mathrm{a}} = \boldsymbol{x}_{k+1}^{\mathrm{b}} + \mathrm{K}_{k+1}\left(\boldsymbol{x}_k^{\mathrm{obs}} - \boldsymbol{x}_{k+1}^{\mathrm{b}}\right)$
- Generate the analysis error covariance matrix: $\mathrm{P}_{k+1}^{\mathrm{a}} = (\mathrm{I} - \mathrm{K}_{k+1})\mathrm{P}_{k+1}^{\mathrm{b}}$

Similarly, for the ensemble Kalman filter the analysis steps are modified and becomes:

- Generate N vectors corresponding to the measurement error: $\boldsymbol{\varepsilon}_{k+1}^{\prime,i} \sim \mathcal{N}(0, \mathrm{R'})$
- Compute the approximate measurement model error covariance:

$$\mathrm{R}_{k+1}^{\prime,\mathrm{N}} = \frac{1}{\mathrm{N}-1}\sum_{i=1}^{\mathrm{N}}\left(\boldsymbol{\varepsilon}_{k+1}^{\prime,i}\right)\left(\boldsymbol{\varepsilon}_{k+1}^{\prime,i}\right)^{\mathrm{T}}$$

- Kalman gain is computed using the general expression:

$$\mathrm{K}_{k+1} = \mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}}\left(\mathrm{P}_{\boldsymbol{y}\boldsymbol{y}_{k+1}}^{\mathrm{b}}\right)^{-1} \tag{VI.21}$$

The matrices $\mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}}$ and $\mathrm{P}_{\boldsymbol{y}\boldsymbol{y}_{k+1}}^{\mathrm{b}}$ are then approximated using the ensemble mean:

$$\mathrm{P}_{\boldsymbol{x}\boldsymbol{y}_{k+1}}^{\mathrm{b}} \approx \frac{1}{\mathrm{N}-1}\sum_{i=1}^{N}\left[\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}-\overline{\boldsymbol{x}_{k+1}^{\mathrm{b}}}\right)\left(\boldsymbol{x}_{k+1}^{\mathrm{b},i}-\overline{\boldsymbol{x}_{k+1}^{\mathrm{b}}}\right)^{\mathrm{T}}\right]$$

$$\mathrm{P}^{\mathrm{b}}_{yy_{k+1}} \approx \frac{1}{\mathrm{N}-1}\sum_{i=1}^{N}\left[\left(\boldsymbol{x}^{\mathrm{b},i}_{k+1} - \overline{\boldsymbol{x}^{\mathrm{b}}_{k+1}}\right)\left(\boldsymbol{x}^{\mathrm{b},i}_{k+1} - \overline{\boldsymbol{x}^{\mathrm{b}}_{k+1}}\right)^{\mathrm{T}}\right] + \mathrm{R}'^{,\mathrm{N}}_{k+1}$$

- At last, forecasted states are analyzed using the following formula:

$$\boldsymbol{x}^{\mathrm{a},i}_{k+1} = \boldsymbol{x}^{\mathrm{b},i}_{k+1} + \mathrm{K}_{k+1}\left(\boldsymbol{x}^{\mathrm{obs}}_{k} + \boldsymbol{\varepsilon}'^{,i}_{k+1} - \boldsymbol{x}^{\mathrm{b}}_{k+1}\right)$$

The analyzed state is given by the ensemble mean: $\overline{\boldsymbol{x}^{\mathrm{a}}_{k+1}} = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} \boldsymbol{x}^{\mathrm{a},i}_{k+1}$

3. Relation between analysis step and LSE

We here consider that the dynamic model is failing in the sense that it gives forecast states that are almost null. To simplify, we assume that the forecast state is equal to the vector null. In addition, the observation model is a linear observation model obtained by LSE. Let us denote $\mathrm{X} \in \mathbb{R}^{n\times\mathrm{N}}$ the data used to train the models (N is the number of time samples used to train the dynamic and observation models), $\mathrm{Y} \in \mathbb{R}^{p\times\mathrm{N}}$ the observation used to train the models. Then the observation model H is given by $\mathrm{H}^{\mathrm{T}} = (\mathrm{X}\mathrm{X}^{\mathrm{T}})^{-1}\mathrm{X}\mathrm{Y}^{\mathrm{T}}$. In addition, since the dynamic model is assumed to give results close to the vector null the matrix P^{b} is close to $\mathrm{X}\mathrm{X}^{\mathrm{T}}$. For the following, we supposed that $\mathrm{P}^{\mathrm{b}} = \mathrm{X}\mathrm{X}^{\mathrm{T}}$.

Also the LSE of X from Y is given by: $\mathrm{X} = \mathrm{K}_{\mathrm{LSE}}\mathrm{Y}$ with $\mathrm{K}^{\mathrm{T}}_{\mathrm{LSE}} = (\mathrm{Y}\mathrm{Y}^{\mathrm{T}})^{-1}\mathrm{Y}\mathrm{X}^{\mathrm{T}}$

Then, with previous hypothesis, the Kalman analysis step is: $\boldsymbol{x}^{\mathrm{a}}_{k+1} = \mathrm{K}_{k+1}\boldsymbol{y}_{k+1}$ with $\mathrm{K}_{k+1} = \mathrm{P}^{\mathrm{b}}_{k+1}\mathrm{H}^{\mathrm{T}}\left(\mathrm{H}\mathrm{P}^{\mathrm{b}}_{k+1}\mathrm{H}^{\mathrm{T}} + \mathrm{R}\right)^{-1}$.

In addition,

$$\mathrm{P}^{\mathrm{b}}_{k+1}\mathrm{H}^{\mathrm{T}} = \mathrm{X}\mathrm{X}^{\mathrm{T}}(\mathrm{X}\mathrm{X}^{\mathrm{T}})^{-1}\mathrm{X}\mathrm{Y}^{\mathrm{T}} = \mathrm{X}\mathrm{Y}^{\mathrm{T}}$$

$$\mathrm{H}\mathrm{P}^{\mathrm{b}}_{k+1}\mathrm{H}^{\mathrm{T}} = \mathrm{Y}\mathrm{X}^{\mathrm{T}}(\mathrm{X}\mathrm{X}^{\mathrm{T}})^{-1}\mathrm{X}\mathrm{Y}^{\mathrm{T}}$$

$$\begin{aligned}\mathrm{R} = (\mathrm{Y}-\mathrm{H}\mathrm{X})(\mathrm{Y}-\mathrm{H}\mathrm{X})^{\mathrm{T}} &= \mathrm{Y}\mathrm{Y}^{\mathrm{T}} - \mathrm{H}\mathrm{X}\mathrm{Y}^{\mathrm{T}} - \mathrm{Y}(\mathrm{H}\mathrm{X})^{\mathrm{T}} + (\mathrm{H}\mathrm{X})(\mathrm{H}\mathrm{X})^{\mathrm{T}} \\ &= \mathrm{Y}\mathrm{Y}^{\mathrm{T}} - \mathrm{Y}\mathrm{X}^{\mathrm{T}}(\mathrm{X}\mathrm{X}^{\mathrm{T}})^{-1}\mathrm{X}\mathrm{Y}^{\mathrm{T}} - \mathrm{Y}\mathrm{X}^{\mathrm{T}}(\mathrm{X}\mathrm{X}^{\mathrm{T}})^{-1}\mathrm{X}\mathrm{Y}^{\mathrm{T}} + \mathrm{H}\mathrm{P}^{\mathrm{b}}_{k+1}\mathrm{H}^{\mathrm{T}} = \mathrm{Y}\mathrm{Y}^{\mathrm{T}} - \mathrm{H}\mathrm{P}^{\mathrm{b}}_{k+1}\mathrm{H}^{\mathrm{T}}\end{aligned}$$

Thus,

$$\mathrm{K}_{k+1} = \mathrm{X}\mathrm{Y}^{\mathrm{T}}(\mathrm{Y}\mathrm{Y}^{\mathrm{T}})^{-1} = \mathrm{K}_{\mathrm{LSE}}$$

Therefore, in these conditions, the analysis step of the Kalman filter is equivalent to the LSE using the observations to estimate the state.

G. Derivation of POD Galerkin model

Additional information on the derivation of POD Galerkin models can be found in Noack et al. [107].

The flow is supposed to be viscous and incompressible. The incompressible Navier-Stokes equation for the momentum conservation is thus (gravity and external forces in general are neglected):

$$N(\mathbf{u}) = \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{u} + \frac{1}{\rho}\boldsymbol{\nabla}\mathrm{P} - \nu\Delta\mathbf{u} = 0$$

The flow velocity $\mathbf{u}$ is decomposed on a POD base: $\mathbf{u}(\boldsymbol{x},t)=\sum_{i=0}^{n}a^{(i)}(t)\boldsymbol{\Phi}^{(i)}(\boldsymbol{x})$ (POD mode 0 corresponds to the average field).

The Navier-Stokes equation is now projected onto each POD mode and $\mathbf{u}$ is substituted by its POD decomposition in $N(\mathbf{u})$ $\left(N(\mathbf{u}),\boldsymbol{\Phi}^{(i)}\right)_\Omega=0$, thus $\forall i\in[\![0,n]\!]$:

$$\left(\frac{\partial\mathbf{u}}{\partial t},\boldsymbol{\Phi}^{(i)}\right)_\Omega=\frac{\mathrm{d}a^{(i)}(t)}{\mathrm{d}t}$$

$$\left((\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{u},\boldsymbol{\Phi}^{(i)}\right)_\Omega=\sum_{j=0}^{n}\sum_{k=0}^{n}\left(\left(\boldsymbol{\Phi}^{(j)}\cdot\boldsymbol{\nabla}\right)\boldsymbol{\Phi}^{(k)},\boldsymbol{\Phi}^{(i)}\right)_\Omega a^{(j)}a^{(k)}=-\sum_{j=0}^{n}\sum_{k=0}^{n}q_{ijk}^{c}a^{(j)}a^{(k)}$$

So, $q_{ijk}^{c}=-\left(\left(\boldsymbol{\Phi}^{(j)}\cdot\boldsymbol{\nabla}\right)\boldsymbol{\Phi}^{(k)},\boldsymbol{\Phi}^{(i)}\right)_\Omega$.

Then,

$$\left(\frac{1}{\rho}\boldsymbol{\nabla}\mathrm{P},\boldsymbol{\Phi}^{(i)}\right)_\Omega=-f_i^p$$

$$\left(-\nu\Delta\mathbf{u},\boldsymbol{\Phi}^{(i)}\right)_\Omega=-\sum_{j=0}^{n}\left(\nu\Delta\boldsymbol{\Phi}^{(j)},\boldsymbol{\Phi}^{(i)}\right)_\Omega a^{(j)}=-\nu\sum_{j=0}^{n}l_{ij}^{\nu}\,a^{(j)}$$

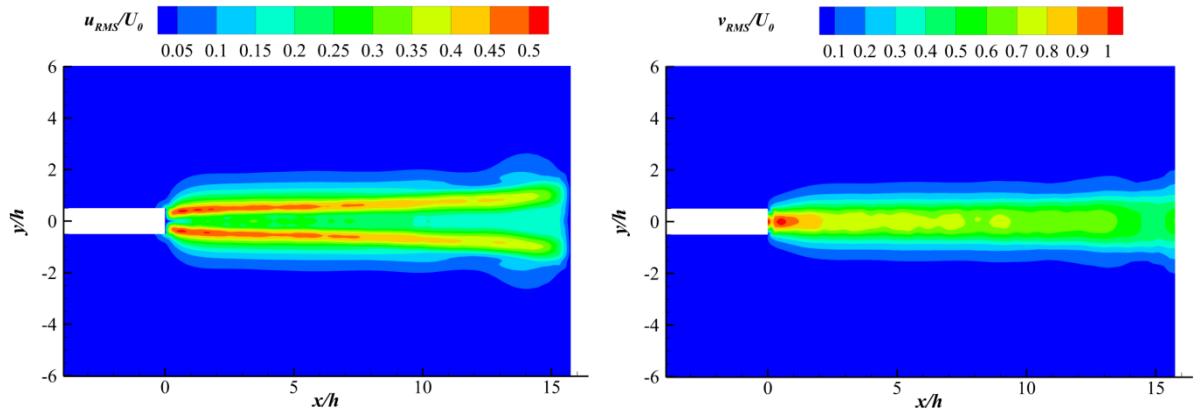
So, $l_{ij}^{\nu}=\left(\Delta\boldsymbol{\Phi}^{(j)},\boldsymbol{\Phi}^{(i)}\right)_\Omega$

Therefore, $\forall i\in[\![0,n]\!]$:

$$\frac{\mathrm{d}a^{(i)}(t)}{\mathrm{d}t}=\nu\sum_{j=0}^{n}l_{ij}^{\nu}\,a^{(j)}(t)+\sum_{j,k=0}^{n}q_{ijk}^{c}\,a^{(j)}(t)a^{(k)}(t)+f_i^p$$

H. Additional results and figures

1. Estimation of the wake downstream of a blunt trailing edge



a) b)

Fig. H.1: a) Streamwise and b) vertical root-mean-square velocities estimated by LSE with 4 pressure sensors (validation set)

2. Estimation of the flow around the OAT15A airfoil (ZDES)

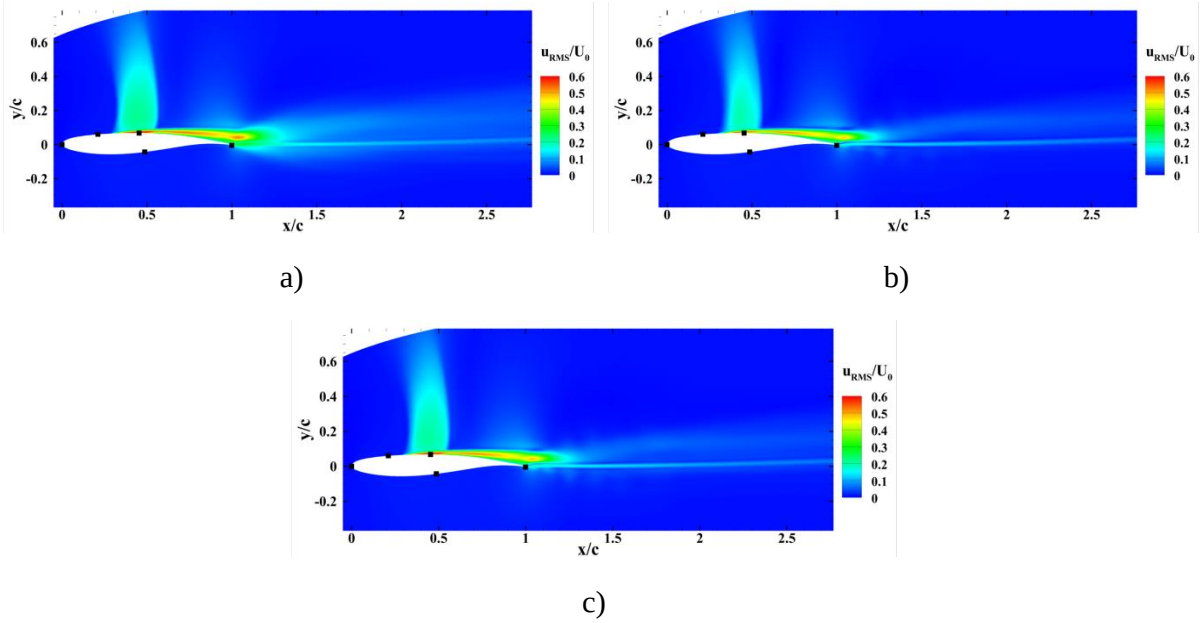


Fig. H.2: Vertical root-mean-square velocity calculated from the validation set, a) original data, b) LSE estimation with “Intuitive P 5” sets of pressure measurement and c) QSE estimation

3. Estimation of the Backward Facing Step flow (low speed PIV database)

Upstream velocity ($m.s^{-1}$)		20	25	30
R^2 from original fields	Training	14.5	12.5	12.5
	Validation	9.7	9.7	9.3

Table H.1: Determination coefficients of the best MTD-LSE situation obtained in 20, 25 and 30 $m.s^{-1}$ of upstream velocity

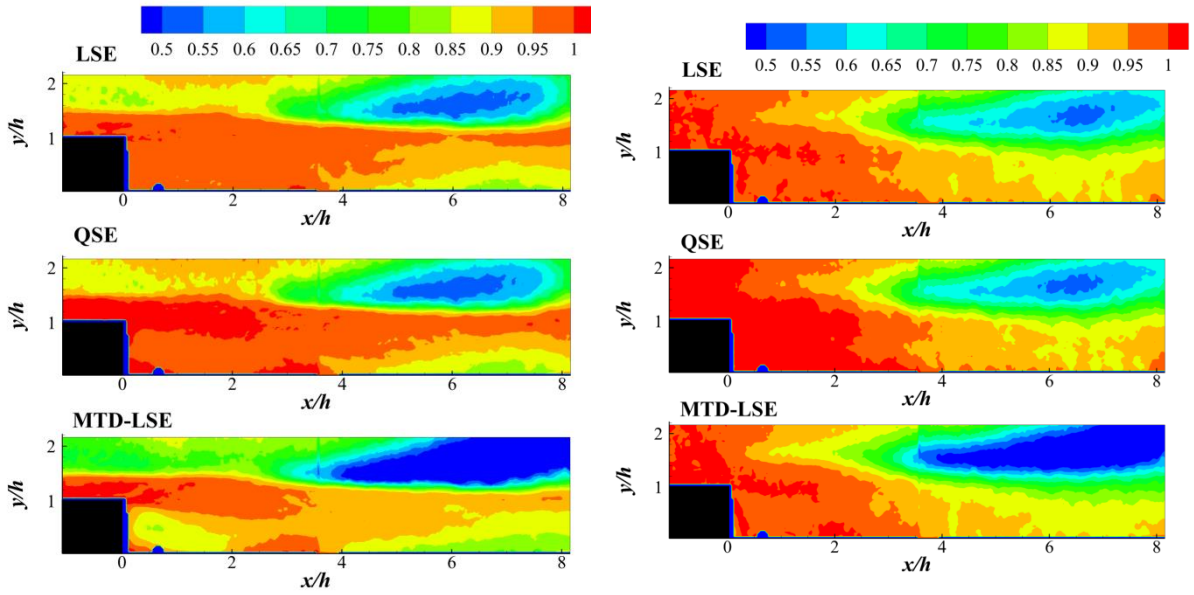


Fig. H.3: Normalized mean square error of the LSE, QSE and best MTD-LSE predictions for the streamwise velocity (a) and the vertical velocity (v) (BFS low speed PIV, 25 $m.s^{-1}$)

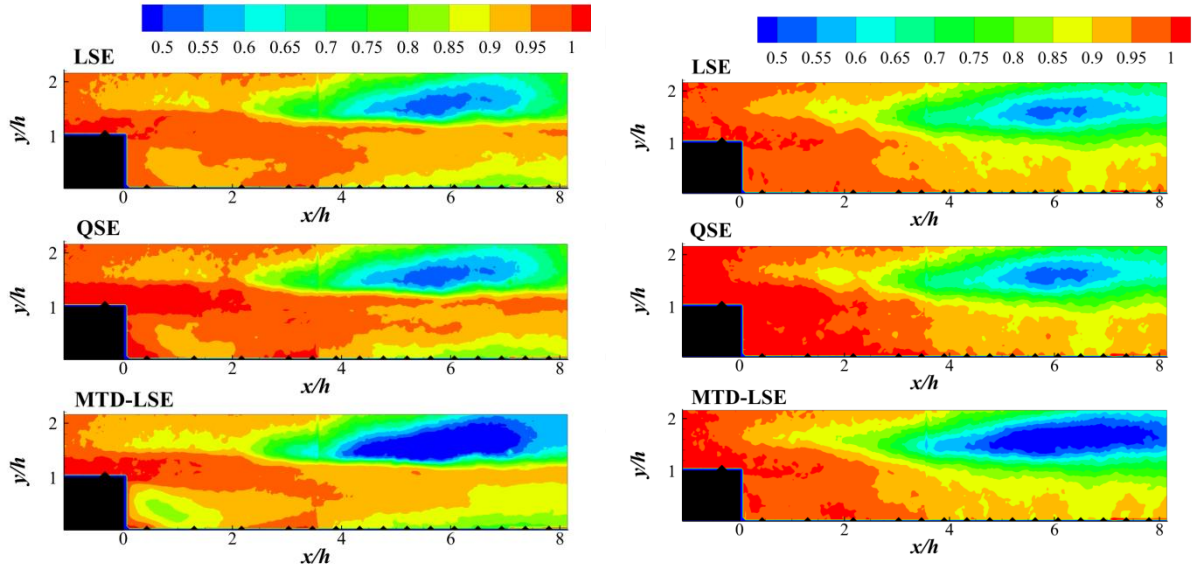


Fig. H.4: Normalized mean square error of the LSE, QSE and best MTD-LSE predictions for the streamwise velocity (a) and the vertical velocity (v) (BFS low speed PIV, 20 $m.s^{-1}$)

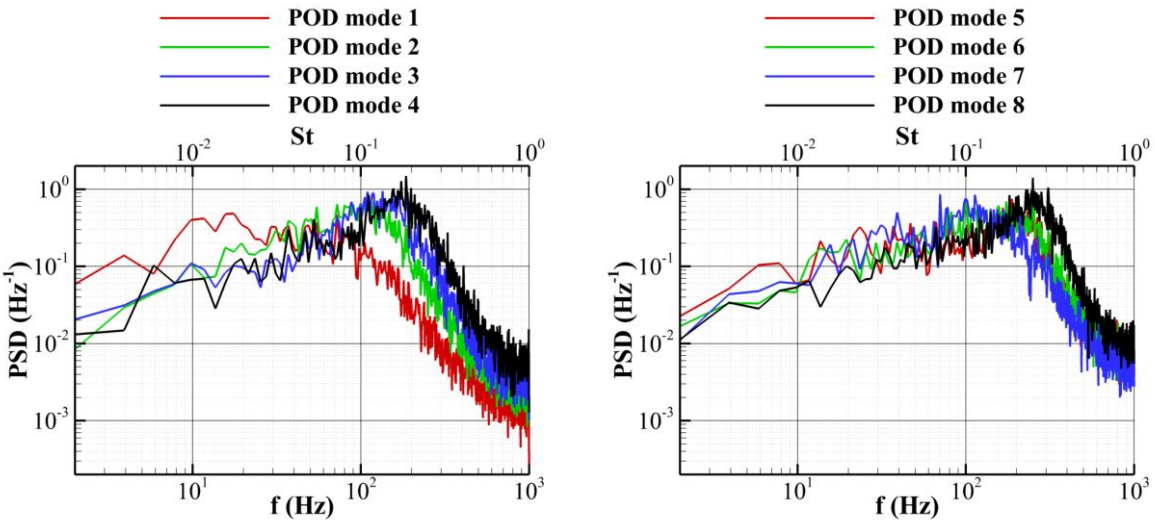


Fig. H.5: Power Spectral Density of POD modes 1 to 4 (a) and 5 to 8 (b) (modes from the low speed PIV database with U0 = 30 m.s-1)

The Fig. H.5 displays the spectra of the first 8 POD modes of the low speed PIV database. These spectra are obtained by projecting the POD modes of the low speed PIV database onto the velocity field of the high speed PIV database in order to obtain the corresponding POD coefficients at 2 kHz and be able to compute their spectra.

4. Estimation of flow around a wall mounted cube

1. Low speed PIV database

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	16.1	22.3	32.5
	Validation	15.7	15	26.3

Table H.2: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: low speed PIV, vertical plane, 10 $m.s^{-1}$; values are expressed in %)

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	21.1	27.1	42.1
	Validation	19.7	19	36.1

Table H.3: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: low speed PIV, horizontal plane, 10 $m.s^{-1}$; values are expressed in %)

		LSE	LSE-POD (90%, 520 modes)	LSE-POD (70%, 94 modes)	LSE-POD (50%, 24 modes)	QSE	Best MTD-LSE
R^2 from original fields	Training	16.7	16.5	16.4	16	23	29.7
	Validation	15.8	15.6	15.6	15.4	14.1	24.5
R^2 from POD projected fields	Training	-	18.4	23.6	31.9	-	-
	Validation	-	17.9	22.9	31.2	-	-
R^2_{POD}	Training	-	1.2	4.8	13.4	-	-
	Validation	-	0.6	4	12.5	-	-

Table H.4: Determination coefficients comparison between LSE, LSE-POD, QSE and MTD-LSE (Cube: low speed PIV, vertical plane, 15 $m.s^{-1}$; values are expressed in %)

		LSE	LSE-POD (90%, 626 modes)	LSE-POD (70%, 104 modes)	LSE-POD (50%, 22 modes)	QSE	Best MTD-LSE
R^2 from original fields	Training	21.4	21.5	21.3	20.7	28	44.7
	Validation	21.2	21.3	21.3	20.8	21.1	35.8
R^2 from POD projected fields	Training	-	23.9	30.4	41.3	-	-
	Validation	-	24.4	30.9	41.7	-	-
R^2_{POD}	Training	-	1.3	5.3	15.6	-	-
	Validation	-	0.9	4.8	15	-	-

Table H.5: Determination coefficients comparison between LSE, LSE-POD, QSE and MTD-LSE (Cube: low speed PIV, horizontal plane, 15 $m.s^{-1}$; values are expressed in %)

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	13.6	19.4	27.3
	Validation	12.4	10.7	20.7

Table H.6: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: low speed PIV, vertical plane, 20 $m.s^{-1}$; values are expressed in %)

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	18.1	24.9	38.6
	Validation	16.5	14.8	29.7

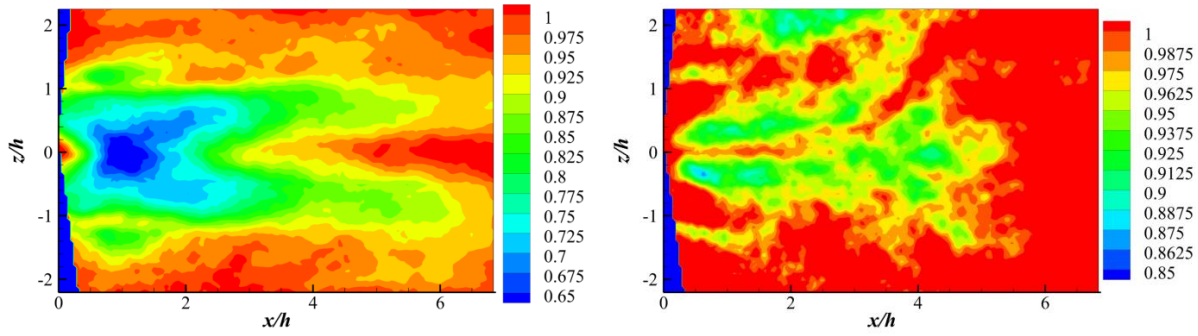
Table H.7: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: low speed PIV, horizontal plane, 20 $m.s^{-1}$; values are expressed in %)

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	14.8	20.9	28.8
	Validation	13.9	12.6	22.6

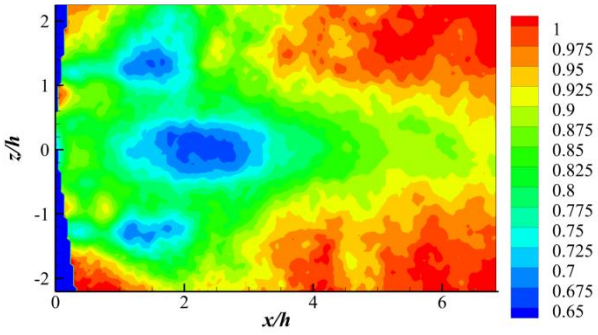
Table H.8: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: low speed PIV, vertical plane, 25 $m.s^{-1}$; values are expressed in %)

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	16.8	23.5	38.8
	Validation	17.2	16.4	31.3

Table H.9: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: low speed PIV, horizontal plane, 25 m.s^{-1}; values are expressed in %)



a) b)



c)

Fig. H.6: Normalized mean square error of the QSE prediction for the streamwise velocity (a), the vertical velocity (b) and the spanwise velocity (c)

2. High speed PIV database

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	17.5	21.6	28.1
	Validation	20.4	23.3	30.1

Table H.10: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: high speed PIV, vertical plane, 10 m.s^{-1}; values are expressed in %)

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	33.5	38.7	51.1
	Validation	33.4	36.6	46.5

Table H.11: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: high speed PIV, horizontal plane, 10 m.s^{-1}; values are expressed in %)

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	16.3	20	27.2
	Validation	17.6	19.5	26.9

Table H.12: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: high speed PIV, vertical plane, 20 $m.s^{-1}$; values are expressed in %)

		LSE	QSE	Best MTD-LSE
R^2 from original fields	Training	23.8	28.5	43
	Validation	24.3	27.2	38.9

Table H.13: Determination coefficients comparison between LSE, QSE and MTD-LSE (Cube: high speed PIV, horizontal plane, 20 $m.s^{-1}$; values are expressed in %)

5. Turbulent integral length scales (Backward facing step low speed PIV)

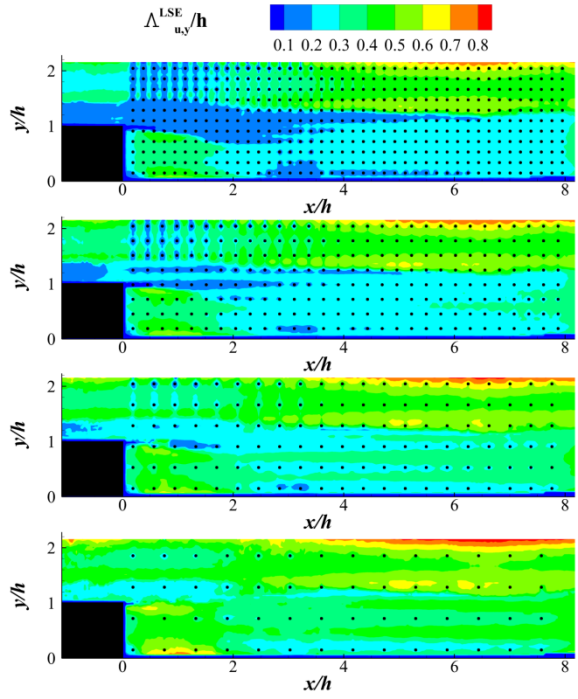


Fig. H.7: Turbulence spatial integral length scales for u calculated in the vertical direction for LSE using several grids of 462, 240, 126 and 56 streamwise velocity sensors (from top to bottom)

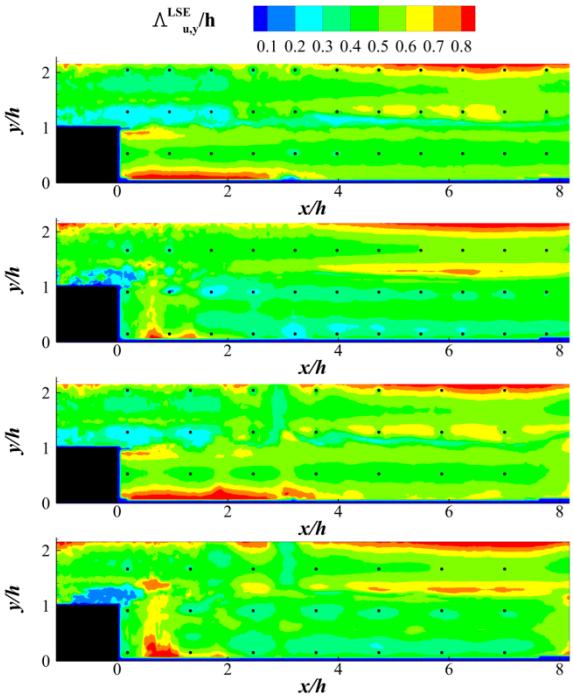


Fig. H.8: Turbulence spatial integral length scales for *u* calculated in the vertical direction for LSE using several grids of 33 and 21 streamwise velocity sensors (from top to bottom)

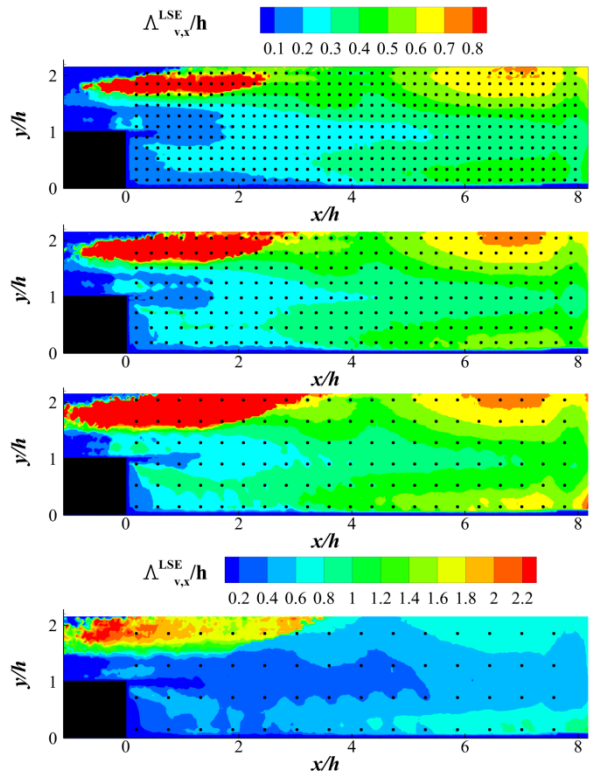


Fig. H.9: Turbulence spatial integral length scales for *v* calculated in the streamwise direction for LSE using several grids of 462, 240, 126 and 56 streamwise velocity sensors (from top to bottom)

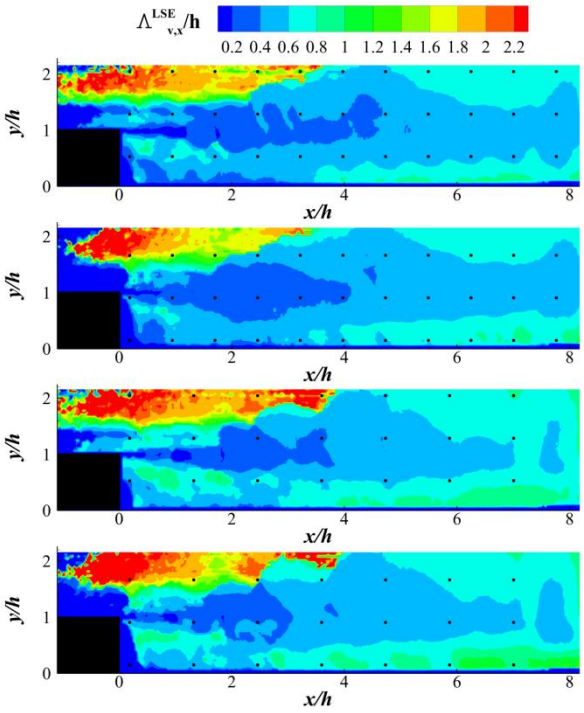


Fig H.10: Turbulence spatial integral length scales for *v* calculated in the streamwise direction for LSE using several grids of 33 and 21 streamwise velocity sensors (from top to bottom)

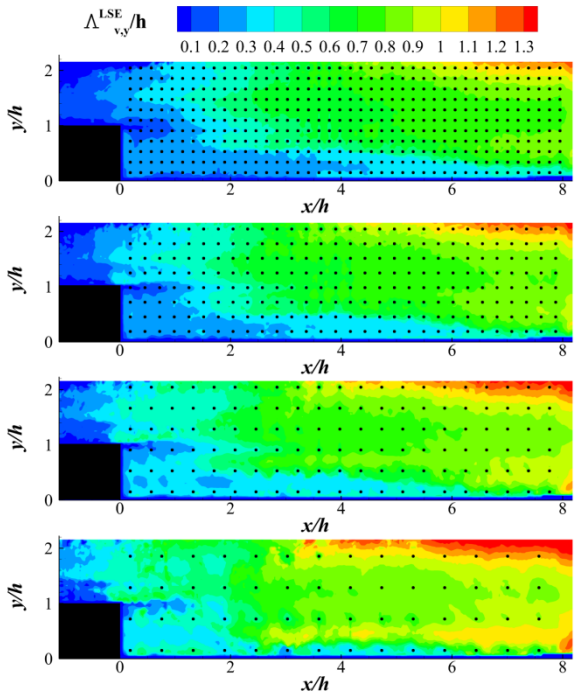


Fig. H.11: Turbulence spatial integral length scales for *v* calculated in the vertical direction for LSE using several grids of 462, 240, 126 and 56 streamwise velocity sensors (from top to bottom)

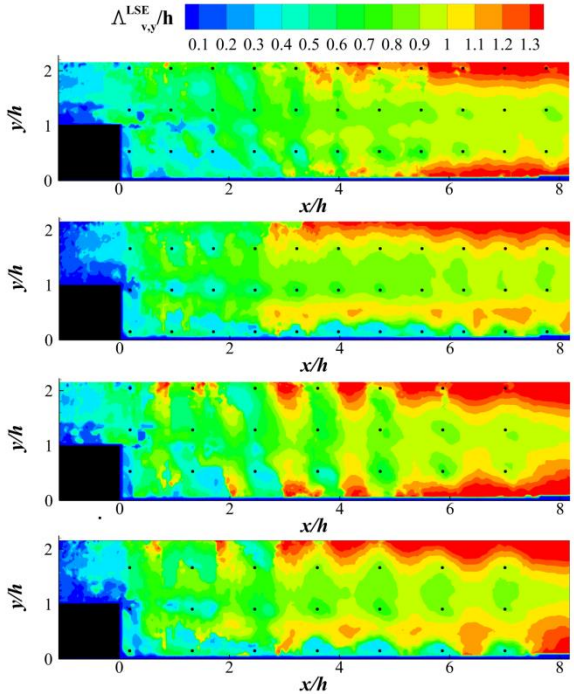


Fig. H.12: Turbulence spatial integral length scales for *v* calculated in the vertical direction for LSE using several grids of 33 and 21 streamwise velocity sensors (from top to bottom)

Reconstruction of turbulent velocity fields from punctual measurements

Real time monitoring of turbulent flows is a challenging task that concerns a large range of applications. Evaluating wake vortices around the approach runway of an airport, in order to optimize the distance between lined-up aircraft, is an example. Another one touches to the broad subject of active flow control. In the aerodynamic field, control of detached flows is an essential issue. Such a control can serve to reduce noise produced by airplanes, or improve their aerodynamic performances. This work aims at investigating and developing tools to produce real time estimation of turbulent velocity fields from a limited number of punctual sensors. After a literature review focused on a popular reconstruction method in fluid mechanics, the Stochastic Estimation (SE), the first step was to evaluate its overall prediction performances on several turbulent flow configurations of gradual complexity. The accuracy of the SE being very limited in some cases, a deeper characterization of the method was performed. The filtering effect of the SE in terms of spatial and temporal content was particularly highlighted. This characterization also pointed out the strong influence of the sensor locations on the estimation quality. Therefore, a sensor location optimization algorithm was proposed and extended to the choice of time delays when using Multi-Time-Delay SE. While using optimized locations for the sensors hold some improvements in accuracy, they were still insufficient for some test cases. The opportunity to use a data assimilation method, the Kalman filter that combines a dynamic model of the flow with sensor information, was investigated. For some cases, the results were very promising and the Kalman filter (and its extensions) outperforms all SE methods.

Keywords: Stochastic Estimation, Kalman filter, regression, data reconstruction, sensor location optimization.

Reconstruction de champs aérodynamiques à partir de mesures ponctuelles

Le suivi en temps réel des écoulements turbulents est une tâche difficile qui trouve des applications dans de nombreux domaines. Un exemple concerne la mesure des tourbillons de sillage au niveau des pistes d'aéroports afin d'optimiser la distance entre les avions en phase d'approche ou de décollage. Un autre exemple se rapporte au contrôle actif des écoulements, en particulier des écoulements détachés, qui représente encore un important défi. De tels contrôles peuvent servir à réduire le bruit des avions, ou bien à améliorer leurs performances aérodynamiques. Cette thèse vise à étudier et à développer des outils afin d'estimer en temps réel des champs de vitesse d'écoulements turbulents à partir d'un nombre limité de mesures ponctuelles. Après une étude bibliographique centrée sur une méthode de reconstruction populaire dans le domaine de la mécanique des fluides, l'estimation stochastique, ses performances sont évaluées pour la prédiction de champs de vitesse issus d'écoulements turbulents de complexité croissante. La précision des estimations obtenues étant très faibles dans certains cas, une analyse plus précise de la méthode est effectuée. Celle-ci a montré en particulier l'effet filtrant de l'estimation stochastique sur le contenu spatial et temporel des champs de vitesse. De plus, le fort impact de la position des capteurs a été mis en avant. C'est pourquoi un algorithme d'optimisation de la position des capteurs est ensuite présenté et également étendu au choix des délais temporels lorsque l'estimation stochastique à délais multiples est utilisée. Bien que l'optimisation de la position des capteurs mène à une amélioration de la précision des prédictions obtenues par estimation stochastique, elle reste néanmoins très faible pour certains cas tests. L'utilisation d'une technique issue du domaine de l'assimilation de données, le filtre de Kalman qui combine un modèle dynamique de l'écoulement avec les mesures, a donc été étudiée. Pour certains écoulements, le filtre de Kalman permet d'obtenir des predictions bien plus précises que l'estimation stochastique.

Mots-clés : Estimation Stochastique, filtre de Kalman, régression, reconstruction de données, optimisation de la position des capteurs.