Unsupervised learning for coherent structure identification in turbulent channel flow

by

John Lyne

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Systems Design Engineering

Waterloo, Ontario, Canada, 2022

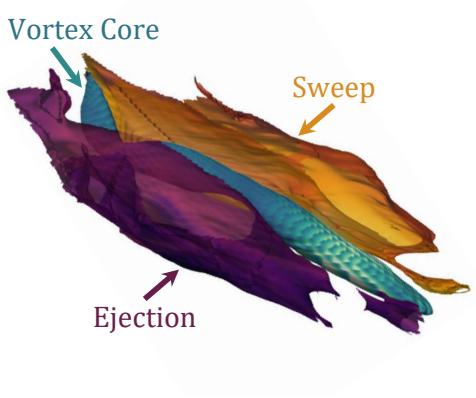
Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Coherent structures (CS), i.e., regions of flow exhibiting significant spatio-temporal coherence, have long been observed in turbulent fluid flow. These CS offer an opportunity to gain insights on fluid behaviour by bypassing the non-linear complexities associated with turbulent flows. Historically, the identification of CS in turbulent flows has involved using manual thresholds to label regions of interest. More recently, work towards more objective threshold selection have used percolation analysis; yet, particular situations can leave the method vulnerable to human bias. This work takes further steps towards pruning human subjectivity from the CS detection process, where an unsupervised learning framework that uses a clustered self-organizing map is used to automatically organize salient regions of flow within a turbulent channel into distinct clusters. The CS identified and analyzed throughout the study include quasi-streamwise coherent vortices, ejections, and sweeps. Structures pertaining to the near-wall region ($y^+ \lessapprox 60$), inner region ($y^+ \lessapprox 100$), and entire wall-normal domain are investigated. Structures are found to agree qualitatively with dynamic expectations, i.e., near-wall vortex structures are quasi-streamwise, and ejection and sweep regions flank vortices. Quadrant distributions of the ejection and sweep structures show larger sweep strength in the lower buffer region ($y^+ \lessapprox 15$) and larger ejection strength above the buffer region ($y^+ \gtrapprox 15$), both characterized by large fluctuating streamwise velocity, whereas streamwise and wall-normal fluctuations in ejections and sweeps that populate the outer layer are more balanced; vorticity component distributions within vortices indicate counter streamwise rotating vortices in the buffer region; orientation statistics of vortices show preference for streamwise orientation in the near-wall, transverse orientation in the log-layer, and no preferred orientation in the outer layer; and the distribution of vorticity transport components, i.e., stretching and tilting, within vortex clusters demonstrate dominant streamwise vortex stretching within buffer layer vortices. Evidence is found of outer layer structures that resemble outsized counterparts of the ejection–vortex–sweep structures found in the near-wall, reinforcing the notion that a hiearchichal self-sustaining process exists in channel flow turbulence.



Acknowledgements

I would like to thank all the people who made this thesis possible. Andrea, for her continued support, guidance, and kindness throughout my graduate studies. Mike and Marek, for reigniting my love for fluid mechanics by way of their wonderful classes, and making mathematics *slightly* less frightening. Amir, for ushering me towards the beauty of physics. The friends I have made at the University, Neil, for his insightful conversations; Javier, for the Spanish lessons; and Zach, for his futile persistence at attempting to surpass me in table tennis. Dominique, for providing a warm light on dark days. My family, for supporting me throughout the pandemic years when attending school was impossible. And to nature, for being a ceaseless source of inspiration and yielding the wondrous dynamics of fluids that have been the subject of my studies.

Dedication

This is dedicated to the ones I love.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Coherent structures in turbulence

The importance of fluids with respect to life on earth is evident. Fluids are vital to the transport of material throughout biological systems, both at the individual and ecosystem level. Many engineered processes rely on leveraging fluid behaviour to exercise specific tasks. Natural selection, the unconscious engineer of life, has created cephalopods, i.e., squid and cuttlefish, that use jet-propulsion to dash around their aquatic environments, and large birds, such as condors and vultures, who conserve energy by using their broad wings to glide from one thermal column to another [Dawkins and Lenzová, 2021, ch. 6]; likewise, humans have consciously engineered machine equivalents: jet-propulsion engines to propel watercraft, and airfoils to lift aircraft. The near-limitless potential of fluid dynamics applications have thus led to considerable efforts to further our understanding of fluid behaviour.

Difficulties associated with developing a firm understanding of fluids lie in the complexities associated with the governing equations of fluid motion. In flow conditions prone to turbulence, ever-increasing spatial ranges of cascading energy transfer and the non-linear contributions to the evolution of fluid motion (e.g., the stretching and tilting of vortices) make causality of physical mechanisms, such as those associated with momentum and energy transfer, difficult to discern. The complexities brought about by this non-linearity act as a shield, fending off physicists from developing a deeper understanding of turbulence physics. Engineers are also affected, e.g., when considering the design of an aerodynamic body, an engineer may contemplate whether to bevel or fillet its edges, or ponder which material to coat its surface with. Having a way to quickly infer how the surrounding fluid, and thus the aerodynamic forces exerted on the body, is affected by these changes saves the engineer time, effort, and resources. One approach to address the

challenges brought about by flow complexity is by analyzing the coherent structures (CS) that arise in turbulent flows, as CS can provide valuable insight into the dynamics of a turbulent fluid system by bypassing some of its complexity. CS are characterized as regions of flow containing spatial and temporal correlations of flow variables, such as velocity fluctuations or vorticity, that persist over a range of length and time scales significantly larger than the smallest local scales [Robinson, 1991]. The visualization of coherent structures has been instrumental to furthering understanding of physical processes in turbulent boundary layers, from the discovery of alternating high and low-speed streaks [Kline and Runstadler, 1960] to horseshoe vortices [Head and Bandyopadhyay, 1981]. Likewise, visualizing CS is also a valuable tool in the design process of surfaces that interact with fluid [Galbraith and Visbal, 2008, Yuan et al., 2018, Hu et al., 2022], providing qualitative insight on how changes in design affect the surrounding flow dynamics.

There currently exist several methods for four-dimensional turbulent CS identification in fluid dynamics simulations. A common approach used is thresholding; for example, Hack and Zaki [2014] visualize low and high speed near-wall streak instabilities in a transitional boundary layer simulation by thresholding the streamwise velocity fluctuations by ±8.5% and plotting the isosurfaces, further visualizing vortical structures by plotting isosurfaces of a vortex identification parameter known as the λ_2 criterion [Jeong and Hussain, 1995] with a threshold of -0.3% and streamwise vorticity fluctuations with a threshold of ±8.5%. In a study of large scale motions in channel flow turbulence, Lee et al. [2014] define low-speed regions of flow by filtering out regions with negative streamwise velocity fluctuations exceeding 10% of the bulk streamwise velocity.

The aforementioned methods, among others [Hedley and Keffer, 1974, Anand et al., 2009, Portwood et al., 2016], require some user-defined value for thresholding. In this situation, a user manually tunes the threshold until the visualization fits perceptual expectations; for example, to visualize the vortical structures in a channel flow with the λ_2 criterion, isosurface threshold values would be adjusted until the expected quasi-streamwise structures are observed. Percolation theory has been used in an effort to move towards more objective threshold selection (see Motoori and Goto [2021], Lozano-Durán et al. [2012]); though, as will be shown, this method does still encounter limitations in particular situations. In this work, a machine learning framework that automatically organizes flow into clusters in an unsupervised fashion is used to address the issues surrounding threshold subjectivity. The boundaries of the identified structures are not bound to a subjective value; instead, these boundaries are automatically learned based on computational pattern recognition.

1.2 Reductionism and holism

Before discussing how computational pattern recognition is used in the field of fluid dynamics, first addressed is the question: Why may the study of coherent structures be valuable?

Turbulent flows are complex systems, and thus furthering understanding of the dynamics of these systems has long involved taking a blend of reductionist and holistic approaches. The philosophical position of reductionism [Sloane, 1945] attempts to describe a complex system as a whole by breaking the system into its constituent parts and traversing up the system hierarchy, using these constituent parts to construct an understanding of the causal mechanisms along the way. Reductionism bases itself on the assumption that a complex system can be described as the sum of its parts. The study of coherent structures in turbulent flows are partially motivated by this perspective, where focusing on local flow patterns and regions of coherence are used to describe global system dynamics. An example being the visualization of coherent vortices to provide insights on the energetics of a fluid simulation. This perspective allows one to filter through the complexity of the system and describe it with a few key features; though, one must be careful not to get lost in the granularity of the system, and diligent about ascribing local features to the entirety of the system.

The opposing perspective of holism (i.e., Gestalt psychology [Koffka, 1935]) lays focus on complex systems as a whole. This perspective lies on the assumption that a complex system is more than the sum of its parts. Considering the system as a whole allows one to view emergent properties of the system that would not observable when looking at individual parts. An example is that of a nervous system, in which a description of the emergence of decision making and memory are not made possible by considering neurons individually, but rather collectively. Holistic analysis allows one to obtain a general description of a system; but, one must be cautious, as it is possible to overlook salient finer details within the system. A relevant holistic example to turbulence research is the use of scalar statistical moments to describe behaviour of a global flow, for example using the root-mean-square (rms) of a velocity fluctuation over the wall direction to describe how energetic the flow is in different wall regions. As a whole, one can observe larger rms quantities in the buffer layer (a highly energetic region in turbulent boundary layers), however, this by no means implies that every region in the buffer layer is highly energetic. Representing a global region's behaviour by a single scalar quantity is bound to gloss over some of the finer details.

It can be tempting to attribute general statistical findings, such as global rms quantities, to an entire population, labeling regions that do not adopt the global trends as low relevance. One must be cautious with this mindset. For example, the cascade of energy in turbulence is generally transferred from larger to smaller scales as slightly larger eddies stretch slightly smaller eddies;

yet, there still exist regions in which energy is transferred against the grain from smaller to larger scales, the process known as backscatter, in which neglecting backscatter in turbulence modeling can lead models to produce inaccurate predictions of perturbation growth [Piomelli et al., 1991]. Likewise, if one only considered the global distribution of streamwise vorticity within a horizontal plane in the buffer layer of a channel flow, they would see a symmetric distribution centered around zero. Though, it is apparent that there exist dynamically salient regions within that plane that are not well-represented by this distribution—counter rotating quasi-streamwise vortices have a bi-modal distribution and are a keystone process to the self-sustainability of channel flow turbulence [Jeong et al., 1997].

Peering out through a foggy window on a gloomy night, to attempt to discern what phenomena are emanating warm light from outside, we would consider not just the faint glimmers of light from the blurry scene; we would instead wipe the window to get a clearer view of the finer details. Nor would we focus solely on regions containing light, we would expand our view to consider the surrounding structures that associate with the light as well. Understanding complex systems requires a blend of both reductionistic and holistic perspectives, which in fluid dynamics involves elucidating the dynamics of fundamental flow components and tying them to the dynamics of the global system. This is what the study of CS in turbulence allows, in which regions of local coherence are observed to elucidate the dynamics of the global domain. Understanding of a system as a whole is necessary to understand how individual components collect to generate emergent properties. Understanding of the fundamental scaffolding is required to shed light on causal mechanisms that give rise to those emergent properties and the overall system dynamics. By studying the CS that arise in turbulent flows, a portion of the system complexity can be bypassed by highlighting local regions of interest, and insights gained through the study of those local features can further our understanding of the global process. This work attempts to refine the way in which CS are identified, through the use of computational pattern recognition.

1.3 Machine learning methods used in turbulence research

Recent advancements in machine learning have ignited the interest to employ machine intelligence across nearly all disciplines. In computational fluid dynamics, of particular interest has been the convolutional neural network (CNN). CNNs work particularly well with structured grid data, first gaining popularity in the computer vision community for working with image data. Simulated flow data shares some structural similarities with that of image data; notably, if taking a two dimensional horizontal slice of velocity vectors in a three-dimensional flow domain, the shape is similar to that of a two dimensional image with three colour channels. Methods developed in computer vision can thus be repurposed to the domain of channel flow turbulence. Supervised

and semi-supervised learning CNN frameworks [Fukami et al., 2019, 2020, Kim et al., 2020] have recently been applied to the super-resolution of turbulent flow fields. Here, super-resolution refers to the process of up-sampling an input from low-resolution to high-resolution. Kim and Lee [2020] also explore the use of unsupervised learning in fluid dynamics, using a generative adversarial network to synthesize temporal sequences of steady, turbulent cross-stream planes of simulated flow velocity fields at various Reynolds numbers.

CNNs clearly possess an array of impressive abilities that enable learning the complex non-linear patterns that arise in fluid dynamics, as shown by the aforementioned studies; however, a limiting factor to take into consideration when employing CNNs is computational memory. CNNs are tailored towards two-dimensional inputs with (typically) 3 colour channels (three-dimensional inputs are used as well for tasks related to video); yet, channel flow simulations contain four-dimensional fields that require significant spatio-temporal resolution. Powerful CNNs that are catered towards image or video inputs already contain millions of learnable parameters, i.e., the neural network weights and biases. The graphics processing units that host and perform the tensor operations associated with neural network training have memory limitations that become increasingly strained when increasing the rank of the input tensors. Hence, Kim et al. [2020] are restricted to using two dimensional slices for their semi-supervised learning framework, and Fukami et al. [2020] are restricted to a low Reynolds number simulation, which requires lower spatio-temporal resolution. In addition to memory restrictions, the CNN architecture itself is excessively complex for the task of identifying CS in channel flow, of which the detection process of CS is simple enough such that the patterns can be found with a simpler neural network configuration, as will be shown.

These drawbacks are mentioned because to identify the spatio-temporal structures in channel flow, a sufficiently large input domain is required to resolve the turbulent structures of interest. While there exist CNNs that perform unsupervised semantic segmentation, the previously noted memory constraints currently make these CNN methods incompatible with four-dimensional channel flow. Worth noting, there do exist some studies that attempt to use machine learning to identify the coherent structures of boundary layer flows. For example, Ströfer et al. [2018] use a CNN to identify the re-circulation region in a two-dimensional flow through a convergent-divergent channel, and identify three-dimensional horseshoe vortices over a wing-body junction. Jagodinski et al. [2020] use a CNN to uncover the near-wall events within a turbulent channel, however, their domain size is significantly restricted and they are required to manually label regions of interest. Both of these studies used a supervised learning scheme, which requires manually labelling ground truths. Having to label regions of interest is sub-optimal, not only because labelling is a cumbersome task, but it also introduces human bias into the learning process, which one would preferably want to avoid. Of preference would be a machine telling the user what the important regions are, not the other way around. This requires not only a method capable

of supporting high-dimensional inputs, but also one that supports an unsupervised learning process. An unsupervised clustering method that satisfies these constraints will be discussed next; however, a brief thought experiment is first provided to familiarize the reader with the concepts of supervised and unsupervised learning.

1.4 Unsupervised learning

Imagine a situation where one observes cloud formations for the first time. Looking up at the sky, observed are some clouds that are large and fluffy, some thin and streaky, some pale white, and some dark grey. Thanks to the development of language and taxonomy, a cloud expert could let one know that the large fluffy clouds belong to the cumulus class, and the thin and streaky clouds belong to the cirrus class. After being taught cloud nomenclature, an assessment could then be set up containing a new set of clouds, where one is asked to assign labels to the new clouds just observed. Upon receiving the results of the test, disappointment attributed to misclassifications, and satisfaction attributed to correct classifications, encourages learning of the correct labels. This situation encapsulates a *supervised* learning task, where the minimization of a loss function with externally labeled data (ground truths) guides learning of the correct labels. However, in the absence of guidance that ground truth labels provide, the brain is yet capable of segmenting and storing these various distinct structures in an *unsupervised* fashion. One still generates internal labels for patterns sensed from external stimuli. Given enough exposure to a wide distribution of clouds, one could formulate internal cloud classifications. This is unsupervised learning: the ability to recognize and learn patterns without ground truth labels.

The goal of this study is to delegate the unsupervised learning process of structure detection to a machine. Rather than observing the structure of clouds in the sky, a machine is tasked with detecting patterns within flow variables provided from direct numerical simulation (DNS) data. The primary unsupervised learning method used in this work is the Self-Organizing Map (SOM) [Kohonen, 1998], a single layer artificial neural network that is trained using a competitive learning scheme. The smaller network size alleviates some of the computational resource allocation pitfalls associated with applying deep neural networks to high-dimensional inputs, allowing for training on high resolution four-dimensional regions of flow. Since its introduction, the SOM has been used to detect patterns in satellite imagery [Richardson et al., 2003]; to model rainfaill-runoff patterns in watershed to aid water resource engineers [Kalteh et al., 2008]; and more recently to cluster a large eddy simulation [Narasimhan et al., 2021] and a DNS [Wu et al., 2019] of a transitional boundary layer into turbulent and non-turbulent regions.

1.5 Thesis outline

This study will use a clustered SOM to identify dynamically important regions in a simulated turbulent wall-bounded flow. First, detailed background on the dynamics and energetics of fluids in wall-bounded flows, as well as the physical structures commonly associated with these dynamics, is provided in § 2. In § 3, details are provided on the simulation data subject to the analyses of the current study, and the data pre-processing required prior to input into the unsupervised learning pipeline. Three scales of the domain will be analysed, and the three pairs of flow variables relevant to these scales will be outlined in this section as well. The methodologies of the learning algorithms, both the SOM and the agglomerative clustering method, and the general machine learning pipeline are outlined in § 4. The unsupervised method will also be compared to standard thresholding and percolation theory in this section. Results are presented in § 5 for each of the three scales, where ejection, sweep, and vortical structures identified by the unsupervised learning method are examined in depth. Resulting structures will also be compared against those identified by percolation theory. Concluding remarks are made in § 6.

Chapter 2

Background

The objective of this study is to highlight the strengths of having an objective method that learns to identify CS within flow fields. This study will focus on two interdependent processes that include: structures relevant to momentum transfer, known as ejections and sweeps; and structures relevant to strong local rotation, known as vortices. First, the fundamental equations that describe fluid motion and the evolution of turbulent kinetic energy (TKE), are presented. Background is then provided on each of the structures and their relevance to the research of boundary layer turbulence.

2.1 Navier-Stokes equations

The motion of fluid can be described by a collection of partial differential equations known as the Navier-Stokes equations. The Navier-Stokes equations for an incompressible, constant density, Newtonian fluid are

$$\frac{\partial u_i}{\partial x_i} = 0, \tag{2.1}$$

$$\frac{Du_i}{Dt} = \frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_j \partial x_j}. \tag{2.2}$$

where p is pressure, ρ is density, and ν is kinematic viscosity. In this work, the x, y, and z axes refer to the streamwise, wall-normal, and spanwise directions, respectively. The velocities in the x, y, and z directions are denoted by u, v, and w, respectively, and $(u_1, u_2, u_3) = (u, v, w)$. These equations represent Newton's laws as applied to a fluid, namely Newton's first law, the conservation of mass (equation 2.1); and Newton's second law, the conservation of momentum

(equation 2.2). Note that the influence of gravity has been absorbed into the pressure term, as outlined in [Wyngaard, 2010, p. 12]. These equations state that the sum of total acceleration of the fluid is equal to the sum of pressure, gravitational, and viscous forces. Equation 2.2 can also be written in dimensionless form

$$\frac{Du_i}{Dt} = \frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{1}{\mathrm{Re}} \frac{\partial^2 u_i}{\partial x_j \partial x_j}, \tag{2.3}$$

where $\mathrm{Re} \sim \frac{uh}{\nu} \equiv \frac{\text{advection}}{\text{viscosity}}$ is the Reynolds number and h is the characteristic length scale of the fluid domain, which, in the case of this study, corresponds to the height of the channel. Highly viscous, slow-moving fluids have a small Re, i.e., $\mathrm{Re} \ll 1$, indicating that advective terms can be neglected, the flow is laminar, and the Navier-Stokes equations approach linearity; fast-moving, low-viscosity fluids have a large Re, i.e., $\mathrm{Re} \gg 1$, indicating the flow of a highly turbulent inviscid fluid. A Reynolds number used to characterize wall-bounded flows is

$$\mathrm{Re}_\tau = \frac{u_\tau h}{\nu} \tag{2.4}$$

[Durbin and Reif, 2010, p. 59] where $u_\tau = (\tau_w/\rho)^{0.5}$ is the friction velocity. Here, τ_w is the wall shear stress. Last, spatial dimensions in channel flow can be scaled into dimensionless quantities [Durbin and Reif, 2010, p. 59], known as wall-unit scaling, e.g., for the wall-unit equivalent of wall-normal position,

$$y^+ = \frac{y u_\tau}{\nu} = \frac{y \mathrm{Re}_\tau}{h}. \tag{2.5}$$

Dimensionless wall-unit scaling enables comparison of analyses across flows with varying Re. Near wall wall-normal regions in boundary layers are the viscous sub-layer (VSL) ($y^+ \lessapprox 5$), the buffer layer ($5 \gtrapprox y^+ \lessapprox 30$), and the logarithmic layer ($60 \gtrapprox y^+ \lessapprox 100$) (figure 2.1) [Kundu, 2003, p. 530].

2.2 Turbulent kinetic energy in wall-bounded flow

Flow within a turbulent boundary layer [Durbin and Reif, 2010, p. 65] undergoes a continual process in which kinetic energy from the free-stream, i.e., mean, flow enhances turbulent fluctuations; meanwhile, kinetic energy dissipates into heat by viscous effects. Herein, a fluctuating quantity refers to a quantity deviating from the mean, i.e., $u' = u - U$, where U represents the mean of u. In the absence of buoyancy effects, the equation to describe the spatio-temporal evolution

of the kinetic energy within turbulent fluctuations, the turbulent kinetic energy (TKE), can be obtained by subtracting the mean momentum

$$\frac{DU_i}{Dt} = \frac{\partial U_i}{\partial t} + U_j \frac{\partial U_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial P}{\partial x_i} + \nu \frac{\partial^2 U_i}{\partial x_j \partial x_j} \tag{2.6}$$

from the total momentum (equation 2.2), multiplying the difference by u_i', and ensemble averaging, yielding

$$\frac{D}{Dt}\left(\underbrace{\frac{1}{2}\overline{u_i'^2}}_{TKE}\right) = -\frac{\partial}{\partial x_j}\left(\underbrace{\frac{1}{\rho}\overline{p'u_j'}}_{\Pi_{ij}} + \underbrace{\frac{1}{2}\overline{u_i'^2 u_j'}}_{T_{ij}} - \underbrace{2\nu\overline{u_i' s_{ij}'}}_{D_{ij}}\right) - \underbrace{\overline{u_i'u_j'}\frac{\partial U_i}{\partial x_j}}_{P_{ij}} - \underbrace{2\nu\overline{s_{ij}'s_{ij}'}}_{\varepsilon_{ij}} \tag{2.7}$$

where

$$s_{ij}' = \frac{1}{2}\left(u_{i,j}' + u_{j,i}'\right) \tag{2.8}$$

is the fluctuating strain rate tensor. TKE production, P_{ij}, appears with opposite sign in the mean kinetic energy (MKE) equation, and is generally negative in boundary layer flow [Durbin and Reif, 2010, p.52; Atoufi et al., 2020, p.14], acting as a sink of kinetic energy from the mean flow and a source of kinetic energy to turbulent fluctuations. Viscous dissipation, ε_{ij}, represents the dissipation of kinetic energy into internal energy, or heat, due to viscous effects. Its contribution to TKE is strictly negative, as $s_{ij}'s_{ij}'$ and ν are strictly positive. Worth noting, energy is also dissipated by the mean flow; however, the scale of mean dissipation is negligible in turbulent flows, of order $\frac{1}{\mathrm{Re}}\frac{u^3}{h}$, compared to the scales of $P \sim \frac{u^3}{h}$ and $\varepsilon \sim \frac{u^3}{h}$. Thus, most MKE is transferred into fluctuating kinetic energy, in which the kinetic energy of the fluctuations cascades to smaller and smaller scales until an eventual dissipation into internal energy, i.e., heat. Turbulent transport and viscous diffusion, T_{ij} and D_{ij}, describe how energy translates throughout the fluid by turbulent dynamics and molecular diffusion, respectively; and Π_{ij} represents the work done by pressure to advect fluctuating flow, an important process that sustains fluid motion in the VSL, where pressure supplements kinetic energy to otherwise slow-moving fluid near a no-slip boundary [Atoufi et al., 2020].

In boundary layer flow, TKE production is most significant in close proximity to the surface of the boundary (just above the VSL) due to large mean wall-normal shear, $\partial U/\partial y \equiv U_{,y}$, where y is the direction perpendicular to the wall and U is mean streamwise velocity. The no-slip boundary condition on the boundary causes a rapid increase in fluid velocity after the VSL (see figure 2.1, or the mean velocity profile in 3.1). This thin, yet dynamically rich, region that encompasses the VSL, buffer layer, and logarithmic layer is known as the near-wall region. The

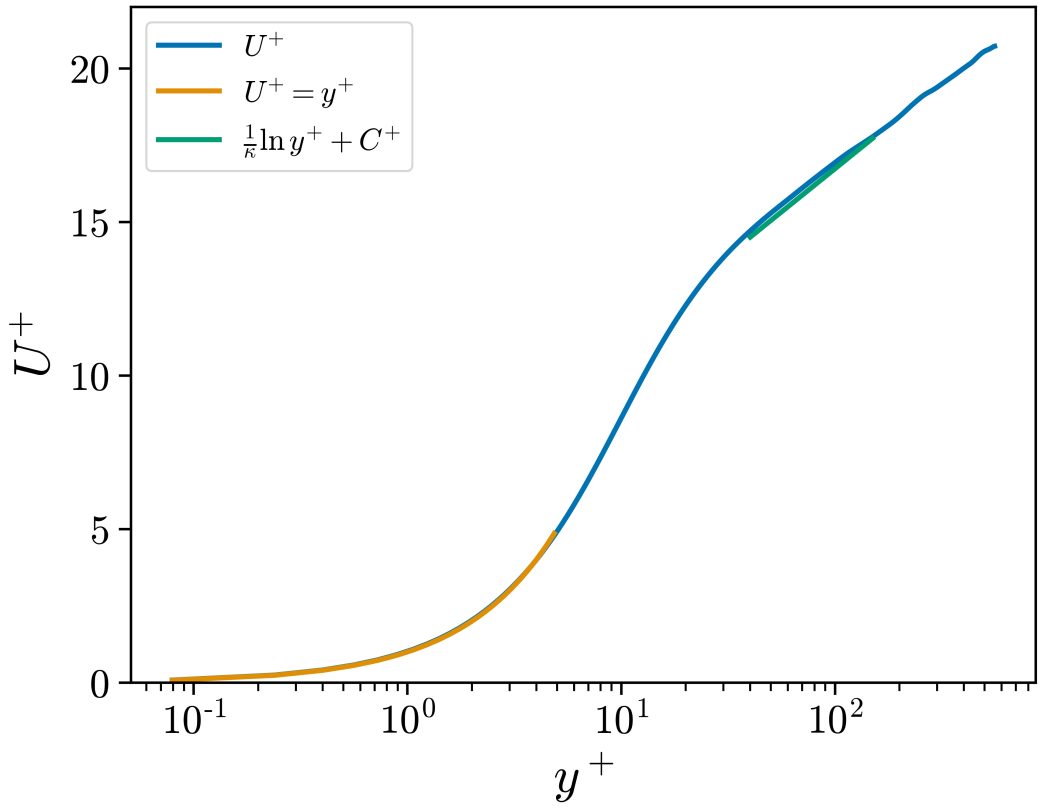


Figure 2.1: Mean velocity profile of a plane channel flow in log-linear coordinates. Yellow: viscous sub-layer; blue region between yellow and green: buffer layer; green: log layer; κ = 2.44; C^+ = 5.5.

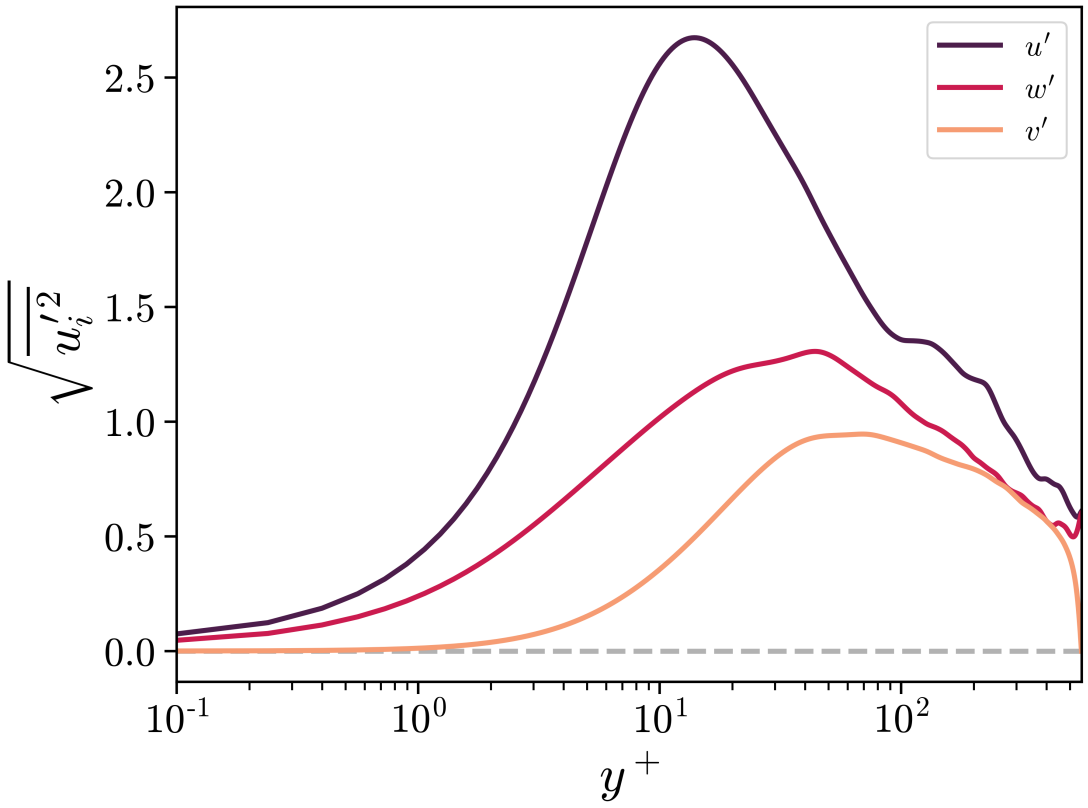


Figure 2.2: Root-mean-square (rms) of fluctuating velocity fields over channel height.

dynamical richness comes in large part from the $P_{12} = -u'v'U_{,y}$ term in equation 2.7, which transfers kinetic energy into streamwise fluctuations that subsequently distribute kinetic energy into lateral velocity fluctuations, to a lesser effect; and wall-normal velocity fluctuations, to an even lesser effect [Atoufi et al., 2020]. As such, the kinetic energy of streamwise fluctuations is generally largest, followed by spanwise fluctuations and then wall-normal fluctuations (see figure 2.2). This energy transfer is most prominent in the buffer layer ($5 \gtrapprox y^+ \lessapprox 30$) where $U_{,y}$ is largest (as per the slope demonstrated in figure 2.1). The question that follows is: How do these energy transfer mechanisms manifest in space and time? Many years of collective boundary layer turbulence research [Head and Bandyopadhyay, 1981, Kim and Moin, 1985, Robinson, 1991, Del Álamo et al., 2006, Jiménez, 2018, Motoori and Goto, 2021] have found that the structures most commonly associated with momentum and energy distribution in shear flows are ejections, sweeps, and quasi-streamwise vortices.

2.3 Ejections, sweeps, and vortices

In the absence of stabilizing effects, the process of turbulence generation and dissipation is self-sustaining [Waleffe, 1997]. As such, understanding the specific mechanisms that transfer energy between the mean and fluctuating flow fields, and dissipate fluctuating energy, have long been sought after by fluid dynamicists. How do these energy extraction and dissipation mechanism manifest in space and time? This is where the study of the internal scaffolding of fluid flow comes in, as quasi-periodic spatio-temporal patterns have long been observed in boundary layer flow [Kline and Runstadler, 1960, Head and Bandyopadhyay, 1981]. These coherent motions provide an unique lens into the otherwise complex dynamics of flow, and investigating them can inform guidance of turbulence control; shed light on the dynamic phenomena that underlie the gross statistics that are measured through modeling; and elucidate causality in fluid interactions.

Fluid motions in shear flows can be sorted into quadrants [Wallace et al., 1972, Willmarth and Lu, 1972] (figure 2.3) relative to their contributions to the turbulent instantaneous Reynolds stress, $-u'v'$. Quadrant splitting is a simple yet intuitive method that involves separating streamwise and wall-normal velocity fluctuations into quadrants based on the sign (i.e., + or -) of the fluctuating velocity component. A positive wall-normal velocity fluctuation ($v' > 0$) paired with a negative streamwise velocity fluctuation ($u' < 0$) is considered to be a part of an ejection, or Q2 event, since the flow moves upwards and against the mean streamwise flow; with the signs switched it is considered a sweep, or Q4 event, since the flow moves downwards and exceeds the mean streamwise flow. By the late twentieth century it was widely agreed upon that fluid dynamics in the wall-region ($y^+ < 100$) contributed significantly to the production of TKE, due to the considerably large mean wall normal shear ($U_{,y}$); and large Reynold stresses, related to ejections

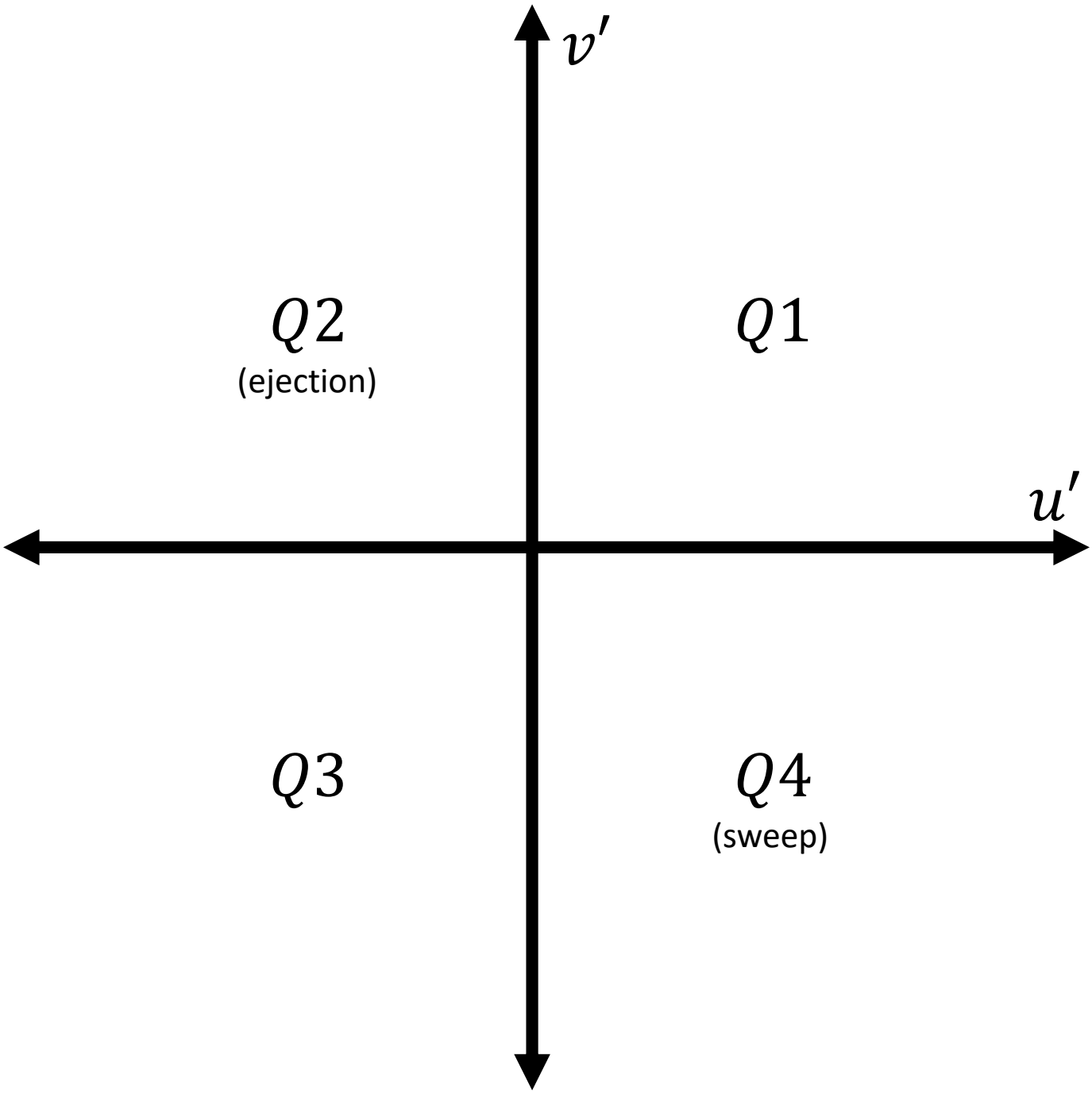


Figure 2.3: Quadrant schematic of the instantaneous Reynolds stress plane.

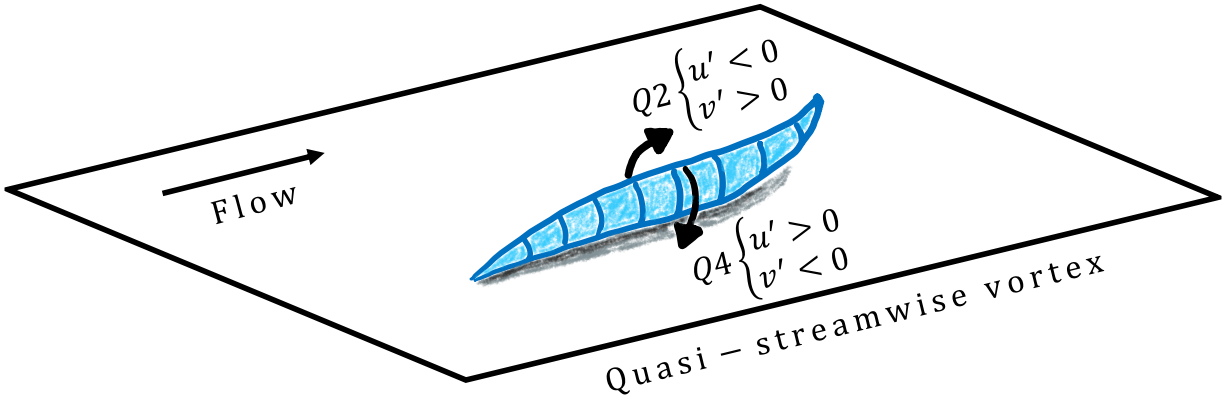


Figure 2.4: Conceptual model of the relationship between quasi-streamwise vortices, ejections, and sweeps in the near-wall region.

and sweeps, found in this region. Furthermore, structures such as near wall streaks; ejections and sweeps; near wall shear layers; quasi-streamwise vortices in the wall region; and transverse vortices in the outer region were deemed of significant importance in terms of understanding the self-sustaining process of channel flow turbulence [see Robinson, 1991, pp. 32-33].

The complexities associated with the Navier-Stokes equations, notably the vast span of length scales required to resolve energy transfer in turbulence, had long made accurately modelling turbulent fluid flow seem impossible. That is until some significant advancements in computing arrived in the late 1980s, which enabled the use of computer models to simulate turbulent fluid behaviour, which require increasingly refined grids when increasing Reynolds numbers [Wyngaard, 2010, p. 305]. Of note, these new models included large eddy simulations (LES) (the underlying concepts of LES are laid out in Lilly [1966]); and direct numerical simulations (DNS) (see Orszag and Patterson [1972] for what is considered the first DNS). The former attempts to accurately model the larger spatial scales and parameterizes small scale behaviour; the latter attempts to accurately model all scales, requiring a more resolved grid which increases computational costs. Kim and Moin [1985] embarked on a detailed analysis of a LES dataset, investigating the structural qualities of vortical motions within the modelled flow, leading them to find statistically significant evidence of hairpin-shaped vortical motions angled roughly 45 degrees from the wall to the streamwise direction. Kim and Moin [1986] later found that upright horseshoe vortices tended to be associated with Q2 events, while inverted horsehoes were more

associated with Q4 events. Shortly thereafter, Robinson [1991] carried out the first in-depth structural investigation of the DNS database of Spalart [1988], which included extensive 2D and 3D visualizations of ejections, sweeps, and vortical structures.

Today, further computational advancements have allowed for increasing in-depth analyses of DNS databases. The structural qualities of vorticity structures [Del Álamo et al., 2006] and momentum-transferring structures [Lozano-Durán et al., 2012, Dong et al., 2017, Jiménez, 2018] have been studied in channel flow DNS at various Reynolds numbers. Large scale motions and very large scale motions [Lee et al., 2014], i.e., regions in the outer layer where flow coheres over exceptional scales of space and time, have had their characteristics anaylsed, as well as how their presence in the outer layer interacts with the dynamics in the wall-region below [Hwang et al., 2016]. The spatial hierarchy of both lifted low streamwise speed structures (i.e., ejections) and rotational structures [Motoori and Goto, 2021] have been studied to further illuminate the physical mechanisms in the energy cascade of turbulent channel flow. This work touches each of these phenomena, i.e., momentum and vortex structures across the entire wall-normal extent of the channel, in which what differentiates this work from the aforementioned is the methodology used to identify regions of interest.

Chapter 3

Data

Prior to discussing the unsupervised learning methodology used to identify turbulent coherent structures, the data to be analysed is first introduced. The velocity fields and velocity gradient fields taken from a DNS of turbulent shear flow [Durbin and Reif, 2010, p. 58] will later be clustered to identify regions of coherence relevant to TKE production and local rotation. The flow configuration used in this section represents the fundamental building block of turbulent fluid flowing over a smooth and solid surface.

3.1 Direct numerical simulation data

A high resolution DNS [Atoufi et al., 2019] of turbulent open channel flow (figure 3.1) is analysed in this paper. The governing equations of this simulation are the incompressible non-dimensional Navier-Stokes equations, equations 2.1 and 2.3. The flow is driven in the streamwise direction by a uniform pressure gradient, and periodic boundary conditions are applied in the horizontal directions. No-slip and no-stress boundary conditions,

$$y = 0 : u = v = w = 0, \tag{3.1}$$

$$y = h : \frac{\partial u}{\partial y} = v = \frac{\partial w}{\partial y} = 0, \tag{3.2}$$

are applied to the bottom and top boundaries, respectively. Fluctuating variables are denoted by a prime (e.g., the fluctuating streamwise velocity is u'), obtained by the subtraction of the horizontal mean. Mean quantities are denoted by a capital letter equivalent of the variable (e.g., mean streamwise velocity is U), and are obtained using averages taken across the horizontal

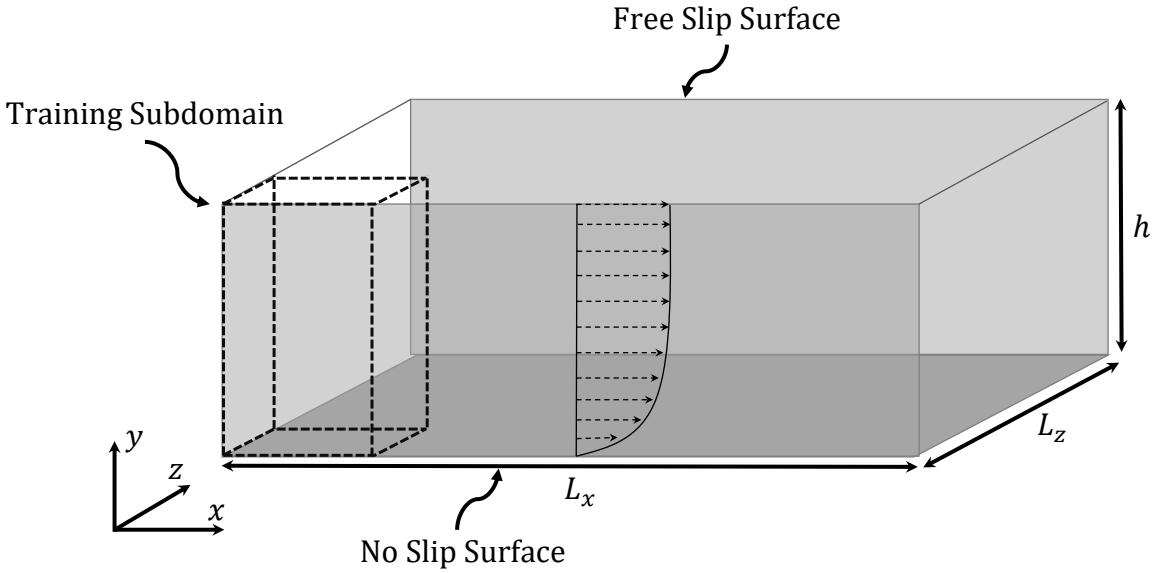


Figure 3.1: Open channel configuration with the mean streamwise velocity profile shown. The sub-domain used for unsupervised model training is also shown.

directions, owing to the horizontal homogeneity of the channel. Note, as the simulation also develops to statistical stationarity (i.e., temporal homogeneity), a temporal average would suffice as well. Wall unit scaled variables are denoted by a '+' superscript (e.g., scaled wall-normal position is y^+), computed using equation 2.5. Further details on the numerical approach of the simulation can be found in Atoufi et al. [2019] and He [2016].

The simulation was performed on a computational domain of size $[L_x/h, L_y/h, L_z/h] = [2\pi, 1, \pi]$ with grid resolution $[N_x, N_y, N_z] = [768, 384, 768]$. Grid spacing is homogeneous in the horizontal directions and non-uniform in the wall-normal directions, with increased resolution near the bottom surface. The friction Reynolds number is $Re_\tau = 560$, a Reynolds number large enough to capture a rich inertial sub-range of TKE transfer [Kundu, 2003, p. 530]. The simulation was run until statistical stationarity with a time step equal to $\Delta t = 0.0002$, and data was recorded at a time interval of $t_s = 0.1$ [Atoufi et al., 2019].

3.2 Flow features of interest

An unsupervised learning method will be used to examine the relationship between regions of significant momentum transfer, i.e., ejections and sweeps, and regions of significant rotation, i.e.,

vortices. The three following subsections will outline the three experiments undergone in this study. Each experiment is distinguished by the particular flow variables that are selected as input to the unsupervised learning framework. The structures identified using these particular flow variables will tend to inhabit particular wall-normal regions in the boundary layer, notably the near-wall region ($y^+ \lessapprox 60$), wall region ($y^+ \lessapprox 100$), and entire wall-normal domain.

3.2.1 Near-wall region

To highlight the dynamics of the near-wall region ($y^+ \lessapprox 60$), vortices are identified using the instantaneous *Q-criterion* [Wray et al., 1988], $Q = 0.5\left(||\boldsymbol{\Omega}||^2 - ||\boldsymbol{S}||^2\right)$ as an input to the SOM, where $\boldsymbol{\Omega}$ and $\boldsymbol{S}$ denote the instantaneous vorticity tensor and strain rate tensor, respectively. The component form derivation for Q can be found in appendix B. To represent regions of significant TKE production and momentum transfer, the dominant instantaneous TKE production term $P_{12} = -u'v'U_{,y}$ is used as input as well, signalling regions in which mean kinetic energy is transferred into TKE. As both Q and P_{12} contain the mean shear, $U_{,y}$, the salient structures relevant to these variables are localized to the near-wall region where the mean shear is maximal.

3.2.2 Wall region

To extend analysis further away from the wall, Q' and the Reynolds stress, $-u'v'$, are instead used, as within these variables is no direct presence of the mean shear. Structures relevant to these variables therefore tend to extend their reach further from the wall. Q' structures are still generally wall-localized ($y^+ \lessapprox 100$), as regardless of the mean shear, gradient terms are stronger near the wall. Reynolds stress structures, lacking gradient terms, generally extend further into the outer layer.

3.2.3 Inner and outer layer ($y \lessapprox h$)

To capture the entire wall-normal extent of the channel, Q' and $-u'v'$ are standardized to zero-mean and unit variance at each channel height, e.g., for Q' we have

$$Q'^s = Q'/\langle Q'^2(y)\rangle^{1/2}. \tag{3.3}$$

Standardization in this way enables a multi-scale analysis of structures across the entire wall-normal extent of the channel. As seen in figure 3.2, strong wall-localized vortical intensities instead get spread near uniformly across the entire wall-normal domain.

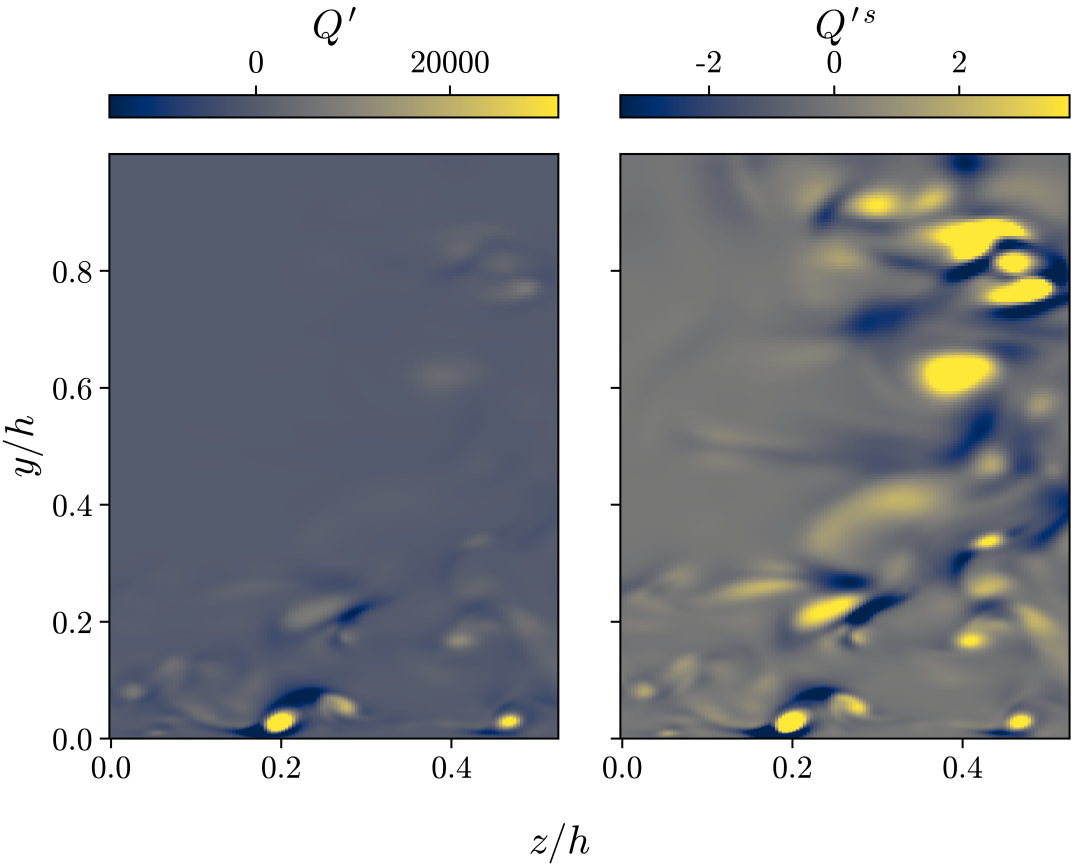


Figure 3.2: Instantaneous cross stream snapshots of Q' (left) and standardized Q' (right), i.e., Q'^s, demonstrating the near wall proclivity of high intensity Q' regions and wall normal ubiquity of high intensity Q'^s regions.

Case	Input Features	N_{rs}	N_v	V_{rs}^{rel}	V_v^{rel}	Section
Near-wall	$-u'v'U_{,y}$, Q	1391	3413	0.5%	0.6%	5.1
Wall	$-u'v'$, Q'	2607	2317	6.0%	0.2%	5.2
Full	$-u'v'^s$, Q'^s	3531	7499	12.4%	2.9%	5.3

Table 3.1: Parameters for each experiment. Each experiment used a sub-domain of shape $[N_t, N_x, N_y, N_z, N_f]_{sub} = [60, 128, 384, 128, 2]$ for training. N_{rs} and N_v are the number of Reynolds stress and vortex clusters identified by the unsupervised learning method, respectively. A distinct structure is considered as a set of highlighted grid cells that connect by grid face, edge, or vertex. V_{rs}^{rel} and V_v^{rel} are the volumes of each population of clusters relative to the full spatio-temporal volume of the sub-domain, e.g., $V_v^{rel} = \sum V_v / (N_t \times N_x \times N_y \times N_z)_{sub}$.

3.3 Data preparation

Prior to model training and the subsequent clustering, the data must be pre-processed. First, a sub-volume of the domain is selected to ease the computational burden on the SOM training process. In the temporal and horizontal dimensions, due to stationarity and homogeneity, the model generalizes well when applied to unseen data in these directions, whereas the full wall-normal extent must be used due to wall-normal inhomogeneity. Care was taken to ensure the temporal and horizontal spatial extents of the sub-domain were greater than the integral time scale and integral length scales, respectively. The sub-domain size selected for model training is $[N_t, N_x, N_y, N_z]_{sub} = [60, 128, 384, 128]$.

With the input flow features of interest selected, the sub-domain data is assembled into $\boldsymbol{D}$ of shape $[N_t, N_x, N_y, N_z, N_f]_{sub}$ where N_f represents the number of features. The spatio-temporal dimensions for each feature are then reshaped into columns to form a feature set, $\boldsymbol{X}$, of shape $[N_t \cdot N_x \cdot N_y \cdot N_z, N_f]_{sub}$. For the near-wall (3.2.1) and wall (3.2.2) sections, prior to being input into the unsupervised framework, each feature is standardized to zero-mean and unit variance, where averages in this case are computed globally, independent of wall-normal position. This step can be interpreted as a means to balance the influence between features during training. For example, the magnitude of Q' tends to be of larger scale than that of $-u'v'$ due to larger gradient terms in the near-wall region, such that in the absence of standardization the model would be more influenced by the larger scale feature during training. To observe inner **and** outer layer (3.2.3) phenomena, standardization is computed as in equation 3.3 using wall-normal position dependent averages. Once standardized, the feature set, $\boldsymbol{X}^s$, is sufficiently prepared for analysis by computational pattern recognition. The methodology of the algorithms used for the pattern recognition task of identifying ejections, sweeps, and vortices will be the topic of the following

section.

Chapter 4

Methodology

Two unsupervised learning algorithms are used in this study to identify regions of interest in a turbulent channel flow simulation. The first algorithm used is the SOM. The choice for this algorithm was noted earlier in section 1.3. The SOM provides a mapping from higher dimensional inputs to a two-dimensional vector space. This two-dimensional space is then clustered using k-means. Reasons as to why the four-dimensional flow fields are not clustered directly will be addressed later in section 4.2. Together, these two algorithms form a framework that allows to deterministically cluster a flow field into regions of interest, which can elucidate coherent structures present within a turbulent flow.

4.1 Self organizing map

The SOM is an unsupervised machine learning method that produces an organized low-dimensional representation of a higher dimensional input. Typically, the low-dimensional representation is in that of a two-dimensional neural map that preserves the topological structure of the data. Throughout training, the map self-organizes such that similarly valued inputs are mapped to similar regions on the map. There exists an iterative and a batch computation method for executing the algorithm. For further details on these two algorithms the reader is referred to Kohonen [2013, pp. 56-58], however brief overviews are provided on both. The iterative method is first introduced to present foundational understanding of the algorithm's objectives, followed by a description of the more powerful batch method used in this work.

4.1.1 Iterative method

Let $\boldsymbol{X}$ denote the input data of shape $[n, f]$, where n is the number of observations and f is the number of features. $\boldsymbol{X}$ is mapped to a neural map of shape $[N, M]$, where N and M correspond to the number of rows and columns in the neural map, respectively. Each node in the neural map has a corresponding weight vector, $\boldsymbol{w}_{ij}$ of shape $[1, f]$, with i and j corresponding to the spatial indices of the neural map. The matrix consisting of these weights is referred to as the weight matrix, $\boldsymbol{W}$, of shape $[N, M, f]$.

In the iterative method, an iteration, t, corresponds to the random selection of one data instance, $\boldsymbol{x}$, of shape $[1, f]$ within $\boldsymbol{X}$. The Euclidean distance between $\boldsymbol{x}$ and all $\boldsymbol{w}_{ij}$,

$$c = \underset{ij}{\operatorname{argmin}}(||\boldsymbol{x}(t) - \boldsymbol{w}_{ij}(t)||), \tag{4.1}$$

is computed, where the node, c, with weights closest to $\boldsymbol{x}$ is termed the best matching unit (BMU) for this input. Nodes that are near the BMU have their weights shifted towards those of the BMU,

$$\boldsymbol{w}_{ij}(t+1) = \boldsymbol{w}_{ij}(t) + h_{c,ij}(t)[\boldsymbol{x}(t) - \boldsymbol{w}_{ij}(t)] \tag{4.2}$$

where $h_{c,ij}$ is the neighbourhood function. In this study, a Gaussian neighbourhood function is used [Kohonen, 1998, p. 2],

$$h_{c,ij}(t) = \exp \frac{-||r_c - r_{ij}||^2}{2\sigma^2(t)}, \tag{4.3}$$

where the numerator represents the Euclidean distance between the BMU, r_c, and other nodes, r_{ij}, on the grid, and the denominator represents a function that monotonically decreases as training progresses. Weights positioned near the BMU, i.e., $||r_c - r_{ij}|| \to 0$, are shifted more, and the radius of influence the Gaussian neighbourhood ($h_{c,ij}$) decreases over time as $\sigma \to 0$, resulting in a neural map with weights that self-organize and localize over time. After sufficient training, all $\boldsymbol{x}$ in $\boldsymbol{X}$ can be mapped to their respective BMU in $\boldsymbol{W}$.

One drawback of the iterative method is that for each iteration, only one random data instance $\boldsymbol{x}$ is selected. For a large spatio-temporal dataset, $\boldsymbol{x}$ represents a tuple of flow variables at a single point in space and time, and our DNS simulation consists of hundreds of millions of these points. Capturing every point during training thus requires hundreds of millions of iterations. In this case, iterating over each individual data instance is intractable and a subset must instead be used; as such, the resulting maps from using the iterative method with large datasets are rarely deterministic, i.e., $\boldsymbol{W}$ between two runs with identical parameters vary due to the random input selection. This is of concern, as for method to bear weight it must provide reproducible outputs. Bussov and Nättilä [2021] demonstrate one workaround to the aforementioned issue, where they

employ multiple iterative SOM method instances to cluster electromagnetic plasma flow fields, and use an ensembling approach to select the best SOM map from the group of instances. Yet, this process is both cumbersome and avoidable, and to avoid the problems surrounding the iterative method this study instead adopts the batch method.

4.1.2 Batch method

The batch method alleviates the aforementioned issues with scalability and reliability of the iterative SOM method. For stable convergence, one assumes that the expected values of $\boldsymbol{w}_i(t+1)$ and $\boldsymbol{w}_i(t)$ be equal as $t \to \infty$. Applying this change to equation 4.2 suggests

$$\mathbb{E}_t\left[h_{c,ij}(t)\left[\boldsymbol{x}(t) - \boldsymbol{w}_{ij}(t)\right]\right] = 0, \forall ij, \tag{4.4}$$

where $\mathbb{E}_t$ represents the expected value over t iterations. Re-writing the expected value as $(1/t)\sum_t(.)$ and isolating $\boldsymbol{w}_{ij}(t)$

$$\boldsymbol{w}_{ij}(t) = \frac{\sum_t h_{c,ij}\boldsymbol{x}(t)}{\sum_t h_{c,ij}}. \tag{4.5}$$

Rather than iterating through random observations one by one, the batch method updates all weights concurrently per update cycle using batches of inputs. One update cycle, t_b, consists of first determining the BMU for each input in a batch, where a batch corresponds to the entire training set, i.e., our entire spatio-temporal set of inputs. Inputs are assigned to the node corresponding to their BMU, and the number of inputs assigned to each node, n_q, are recorded. The mean of the inputs lying within each node, $\overline{\boldsymbol{X}_q}$, are computed, and the weights at each node are updated

$$\boldsymbol{w}_{ij}(t_b) = \frac{\sum_q n_q h_{q,ij}(t_b)\overline{\boldsymbol{X}_q}}{\sum_q n_q h_{q,ij}(t_b)}. \tag{4.6}$$

To summarize, the batch computation follows the process outlined below:

1. For each weight, $\boldsymbol{w}_q$, first compute the mean of all inputs, $\overline{\boldsymbol{X}_q}$, that are nearest to each weight. Record the number of inputs, n_q, falling into each node.

2. For each weight $\boldsymbol{w}_{ij}$, compute the neighbourhood functions, $h_{q,ij}$, between node ij and $q = 1, ..., N \times M$. Update each weight value with equation 4.6.

3. Repeat steps 1 and 2 until $||\boldsymbol{w}_{ij}(t_b) - \boldsymbol{w}_{ij}(t_b - 1)|| < \epsilon$, where ϵ is some small value.

As training progresses the neighbourhood function radius of influence decreases, and once the weight updates settle and become steady the model is considered converged. Batch training provides more consistent and reproducible outputs than the iterative method. A neural map of shape $[N, M] = [10, 10]$ is used in this work, providing a sufficient balance between topological resolution and computational loads. It was found that smaller maps limit topological complexity, important to the clustering step outlined in section 4.2; larger maps were found to increase computational loads with negligible benefit, see appendix B.1.1 for more details. Linear initialization [Kohonen, 2001, p. 107] of weights is employed to accelerate model convergence. Linear initialization involves determining the two eigenvectors, i.e., the principal components, of the autocorrelation matrix of $\boldsymbol{X}$, and spanning these eigenvectors across a two-dimensional linear sub-space. The initial values of $w_{ij}(0)$ are then initialized with this array of points. Note, the autocorrelation matrix of the sub-volumes corresponding to **five** random timesteps was used in this study, as obtaining the principal components of the full sub-domain was too expensive. This is approximation is sufficient, as the rationale around using linear initialization is that any ordered initial approximation should be more profitable than random initialization [Kohonen, 2001, p. 107]. The model took approximately 14 hours to train on two Intel Platinum 8160F Skylake CPUs.

4.2 Clustering

Once trained, the batch SOM method produces a low-dimensional representation of the input data, in the form of $\boldsymbol{W}$. To cluster the data, one option would be to consider each map node as an individual cluster. Wu et al. [2019] follow this approach, using a $[N, M] = [1, 2]$ SOM to identify turbulent and non-turbulent regions in a transitional boundary layer flow, treating instances mapped to one node as turbulent, and others as non-turbulent. In the case of a large neural map this approach leads to a large number of clusters; moreover, it imposes a specific count of clusters, which corresponds to the number of nodes on the map, yet in most cases the true number of clusters is not necessarily known.

Nodes near one another in $\boldsymbol{W}$ hold similarity to one another; hence, it is expected that neighbourhoods of multiple nodes coalesce to form self similar clusters. Therefore, to obtain a distinct cluster set for $\boldsymbol{X}$, $\boldsymbol{W}$ is further partitioned into distinct regions based on converged weight values. Vesanto and Alhoniemi [2000] show the effectiveness of this approach, taking a high-dimensional input and employing this two-level approach to clustering—by first obtaining a SOM map and further clustering the map using an agglomerative clustering method. Applying

the clustering methods directly to large datasets is computationally expensive, whereas clustering the low-dimensional SOM representations of the data instead serves to bypass a large portion of the computational load while maintaining cluster quality.

A similar approach is used here, where the agglomerative clustering method of k-means [Lloyd, 1982] is used to further cluster the SOM. Outputs of k-means are stochastic, sensitive to initialized centroid values; meaning, clusters produced by two separate models trained with identical parameters are not necessarily the same.

4.2.1 Ensuring reproducibility

To obtain a deterministic model that generates reproducible outputs, multiple k-means model instances, $\boldsymbol{m}$, are trained to produce a set of models, $\boldsymbol{m}_{all}$. From this, the best model is selected, $\boldsymbol{m}_{best}$, such that it returns the minimum intra-class variance,

$$\boldsymbol{m}_{best} = \min_{\forall \boldsymbol{m} \in \boldsymbol{m}_{all}} \sum_{b=1}^{k} \sum_{a=1}^{N \times M} ||\boldsymbol{w}_a^{(b)} - \boldsymbol{c}_b||^2, \tag{4.7}$$

where $N \times M$ is the number of data instances (in this case number of SOM nodes), $\boldsymbol{w}_a^{(b)}$ corresponds to a weight instance (in this case the weight vector at node a) falling into cluster b, k corresponds to the number of clusters, and $\boldsymbol{c}_b$ is the centroid value of cluster b. Selecting the best of $O(10^3)$ model instances provides reproducible cluster outputs (see appendix B.1.2), and due to the low dimensionality of the SOM map the computational cost of running many instances is negligible.

Another required input parameter is the number of clusters, k; however, the number of clusters is not necessarily known—injecting that bias into the learning process is to be avoided. To automatically determine the number of clusters, k_{best}, the best models are obtained for cluster counts $k = 2, 3, ..., k_{max}$, setting $k_{max} = 20$ in this study. Two criteria are used to assess the cluster quality of these models: the silhouette coefficient [Rousseeuw, 1987] and the intra-class variance. These two criteria are used to provide robustness to the cluster count determination process. The silhouette coefficient, ranging from $[-1, 1]$, measures how similar data points are to their assigned cluster relative to other clusters, with large positive values indicating strong intra-cluster cohesion and inter-cluster separation; whereas a large negative value indicates that instances are not well-matched to their own clusters and better-matched to neighbouring clusters. Silhouette coefficients are obtained for each cluster count, and the best cluster count relevant to this criteria is that with a maximal silhouette coefficient. For recent studies making use of this criteria see Yatsunenko et al. [2012], Bagirov et al. [2023]. Meanwhile, the intra-class variance for each cluster count is also obtained, and the best cluster count relevant to this criteria is

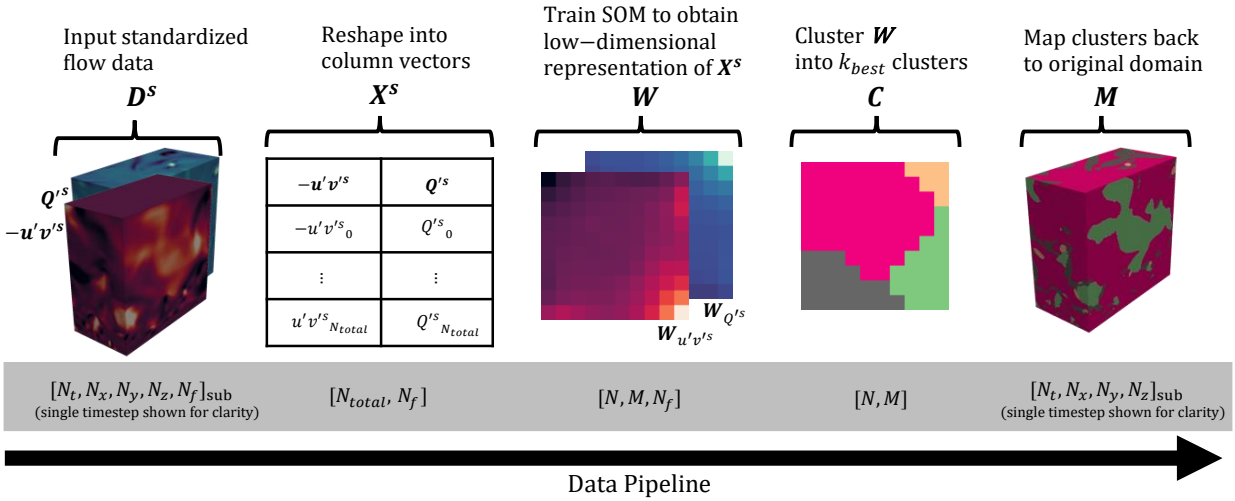


Figure 4.1: Data pipeline for an input containing $N_f = 2$ flow features. Text in grey boxes represent the shape of the data object shown at that point in the pipeline. N_{total} represents the product of the spatial and temporal sizes.

determined using the elbow method [Thorndike, 1953], i.e., the point of the intra-class variance vs. cluster count plot with maximum curvature. For recent studies using this criteria see Omar et al. [2020], Topór [2020]. The average between the two cluster counts, that found by using the silhouette coefficient and that found by the elbow method, is used as k_{best}. For further details on this process, see appendix B.1.3.

4.3 Learning pipeline

Once $\boldsymbol{W}$ is clustered, clusters $\boldsymbol{C} = \{\boldsymbol{C}_1, ..., \boldsymbol{C}_{k_{best}}\}$ are mapped back to the spatio-temporal domain. Using the learned connection between observations and their corresponding BMU on $\boldsymbol{W}$, a mask tensor $\boldsymbol{M}$ of shape $[N_t, N_x, N_y, N_z]$ is created. $\boldsymbol{M}$ is representative of all features that were input in the first stage of the data pipeline, thus having only one feature. In summary, $\boldsymbol{M}$ represents the proper superset of all identified disjoint clusters

$$\boldsymbol{M} \supset \left\{\boldsymbol{C}_1, ..., \boldsymbol{C}_{k_{best}}\right\} \mid \boldsymbol{C}_i \cap \boldsymbol{C}_{j \neq i} = \emptyset. \tag{4.8}$$

An overview of the data pipeline described throughout this section is shown in figure 4.1. Analyses of the learned clusters within $\boldsymbol{M}$ will be the topic of discussion in the section 5.

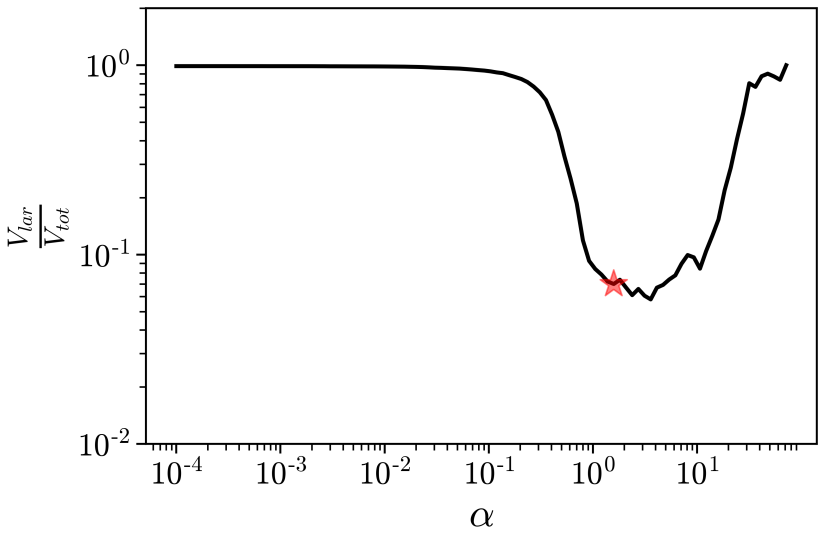


Figure 4.2: Percolation plot for structures corresponding to $Q'^s > \alpha$, ⋆ indicates $\alpha_{perc} \simeq 1.6$.

4.4 Percolation Theory

Prior to the analysis of the clusters produced by the unsupervised learning framework, the methodology of percolation theory is first outlined briefly here. Later in section 5.4, the resulting clusters produced by the unsupervised method will be compared against those found using percolation theory, a method used to determine more objective thresholds to identify coherent structures.

Visualization of CS in numerically represented flows generally involves using judgment to manually select a threshold value that displays the salient regions of the thresholded variable. The threshold tuning process in this case would involves a situation in which the user selects a threshold that brings rise to structures that match their perceptual expectations.

Recent works [Lozano-Durán et al., 2012, Dong et al., 2017, Motoori and Goto, 2021] have used percolation theory to move towards more objective threshold selection. Percolation theory involves first computing the volume of each distinct structure identified by a threshold value. These structures are obtained by simply highlighting regions of flow above the chosen α, and discarding the rest. A distinct structure is considered as a set of highlighted grid cells that connect by grid face, edge, or vertex. The volume of the largest structure, V_{lar}, is compared against the total volume occupied by all structures, V_{tot}, yielding a ratio of V_{lar}/V_{tot}. Ratios are obtained for an array of thresholds, yielding a plot such as figure 4.2 (this plot will be used

later in section 5.4). Note, for small and large α, the percolation plot approaches a ratio of unity. To understand this behaviour, consider that an infinitesimally small threshold selects the entire domain as the structure, thus $V_{lar} \equiv V_{tot}$; and the largest possible threshold corresponds to the global maximum of the thresholded value, thus only highlighting only a solitary "speck" within the domain, resulting again in $V_{lar} \equiv V_{tot}$. A local minimum after a steep gradient on the percolation plot indicates a rapid shift from a coalescence to a distinct and separated set of structures, the threshold value at this minimum is selected to represent the percolation threshold, α_{perc}, which provides a more data-driven threshold than one selected based on visual perception.

Computational analysis was done on Python. The Somoclu Python package (see Wittek et al. [2017]) was used to employ the batch SOM method, scikit-learn was used to employ K-Means clustering, and the PyVista package (see Sullivan and Kaszynski [2019]) was used for three-dimensional structure visualizations.

Chapter 5

Results

Ejections, sweeps, and quasi-streamwise vortices are known to be prevalent in the wall region of wall-bounded flows [Robinson, 1991]. Through the transfer of MKE into TKE, velocity fluctuations are amplified considerably in the near-wall region, primarily by way of mean shear. Momentum transfer in this region is thus enhanced. Characteristic processes involved in momentum transfer include regions of high vorticity with quasi-streamwise orientation, flanked by regions of slower than average (i.e., $u' \ll 0$) fluid being ejected upwards and regions of faster than average (i.e., $u' \gg 0$) fluid being swept downwards Robinson [1991, p. 625]. The presence of these near-wall coherent structures have been long documented in the literature, and more recent works [Dong et al., 2017, Jiménez, 2018] have reaffirmed their presence. In this section, the unsupervised learning method is used to further inspect these processes.

This collection of processes are explored in three stages. First, in subsection 5.1, instantaneous P_{12} and Q fields are used to illuminate the near-wall ($y^+ < 60$) dynamics between vortices and TKE production. In subsection 5.2, the instantaneous $-u'v'$ and Q' fields are used to illuminate the interactions between momentum and fluctuating vorticity CS in the wall region ($y^+ < 100$). In subsection 5.3, the standardized instantaneous $-u'v'^s$ and Q'^s are used to explore momentum and fluctuating vorticity structure interactions across the entire span of the wall-normal domain.

This section will feature analysis done on regions of interest as identified by the unsupervised learning method, involving both qualitative and quantitative methods of analysis. Notably, the quantitative methods will involve evaluating local statistics within the regions of interest. Refining the statistics to intelligently localized phenomena offers the ability to further understand the physical processes that underlay the local phenomena, and how it contributes to the global system.

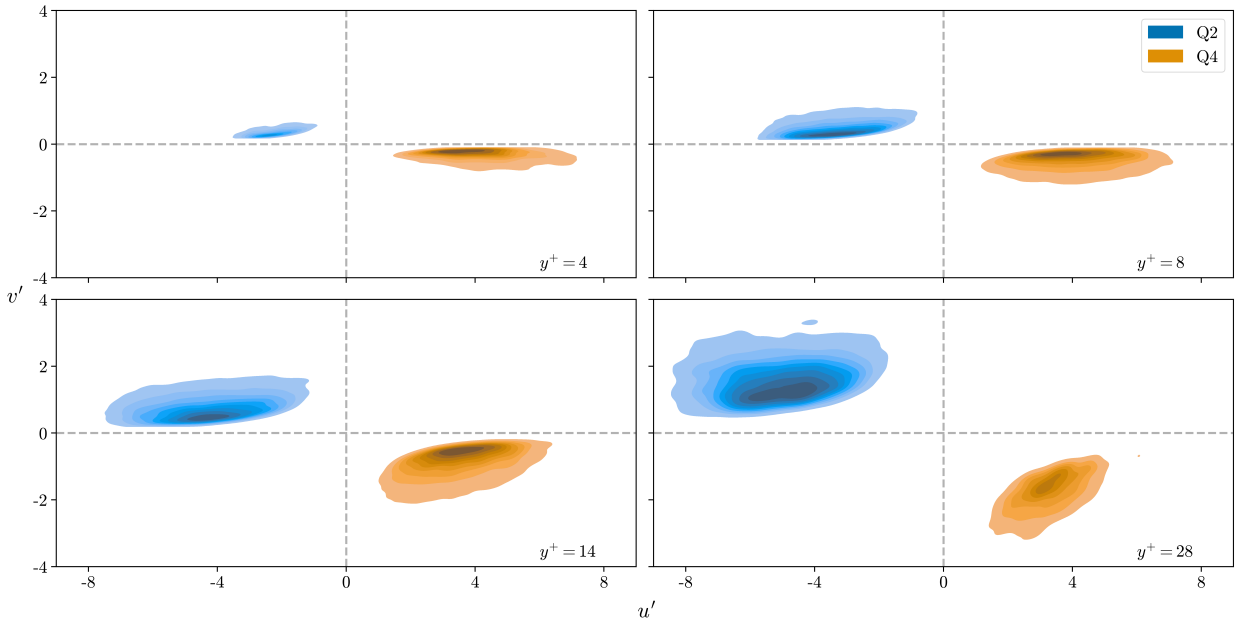


Figure 5.1: Joint $p.d.f(u', v')$ of regions of flow labeled as $C1_{Q2}$ and $C1_{Q4}$ from the a) viscous sublayer, b) lower buffer layer, c) mid buffer layer, and d) upper buffer layer. Contours contain 90% to 10% of the joint probability mass from the lowest to highest contour levels, respectively.

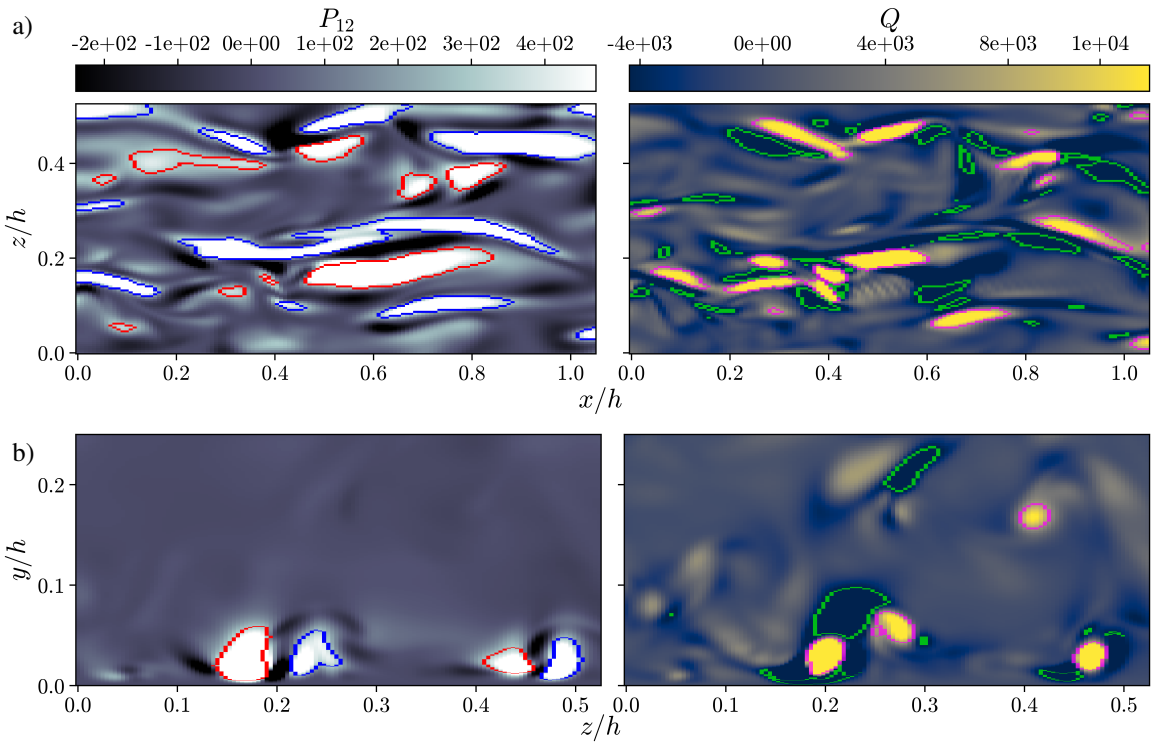


Figure 5.2: Instantaneous a) top-down ($y^+ \approx 15$) and b) cross-stream snapshots of P_{12} (left) and Q (right) overlaid with cluster boundaries. Red ($C1_{Q4}$), blue ($C1_{Q2}$), pink ($C1_{\Omega}$) and green ($C1_S$).

5.1 Near-wall dynamics

The input data for the near wall region analysis, $\boldsymbol{D}_{nw}$, containing fields of instantaneous P_{12} and Q, progresses through the data pipeline shown in figure 4.1. The mask set, $\boldsymbol{M}_{nw}$, is generated, containing $k_{best} = 4$ clusters, forming the cluster set $\boldsymbol{C1} = \{C1_1, ..., C1_4\}$. To link each cluster with its respective dynamic process, quantitative and qualitative methods are employed to further inspect the clusters. The statistics of each cluster with respect to the input variables are first examined to apply meaningful labels.

Cluster $C1_{ns}$ occupies 98% of the channel volume, encompassing what is named as the non-salient region of the channel. While this region is undoubtedly still relevant to the overall dynamics of the channel, it occupies space where production and vorticity/strain are locally low in magnitude. Two clusters, $C1_{\Omega}$ and $C1_S$, are characterized by vorticity dominated regions ($Q \gg 0$)

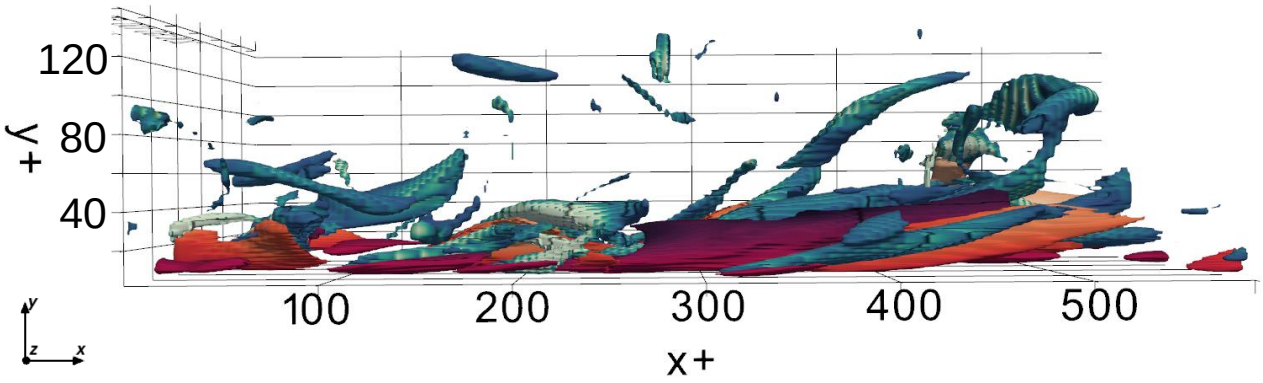


Figure 5.3: Side profile of three-dimensional $C1_\Omega$ (blue) and $C1_P$ (red) structures, showing production structures confined to the buffer layer, and vortices extending out further above. $C1_{ns}$ and $C1_S$ are made transparent.

and strain dominated regions ($Q \ll 0$), respectively. The fourth cluster, $C1_P$, is characterized by regions of large production, altogether forming the cluster set $\boldsymbol{C1} = \{C1_{ns}, C1_\Omega, C1_S, C_P\}$. Figure 5.1 shows that u' and v' values within cells labeled as $C1_P$ are distributed neatly within quadrants Q2 and Q4, which are henceforth denoted by $C1_{Q2}$ and $C1_{Q4}$. Herein, a flow variable within a specific cluster are represented by a superscript indicating the specific cluster (i.e., u'^{C1_Ω} are streamwise fluctuations in the set of grid cells labeled as $C1_\Omega$), akin to a filtered quantity.

Figure 5.1 shows that sweeps within high production regions, i.e., $C1_{Q4}$, are more intense near the wall in the viscous sublayer and early buffer layer. In these regions, the joint distributions demonstrate limited wall-normal fluctuations, as vertical flow is inhibited by the wall below. As wall-normal distance increases, one observes increasing ejection intensities, as well as reduced wall-normal fluctuation restriction. This dichotomy of sweep and ejection intensity in the lower and upper buffer layers is well documented [Wallace, 2016]. Regardless of wall-normal position within the near wall, streamwise fluctuations are generally stronger than wall normal fluctuations, one reason being that the mean shear injects kinetic energy directly into streamwise fluctuating kinetic energy.

The $i = 1$ and $j = 2$ production term from equation 2.7 is large due to the significantly large mean shear, $U_{1,2}$ (or equivalently $\frac{\partial U}{\partial y}$), transferring MKE into streamwise fluctuating kinetic energy. Figure 5.2 and 5.4 show that significant TKE production occurs on the flank of vortices, as shown by the overlaid $C1_{Q2}$, $C1_{Q4}$, and $C1_\Omega$ clusters. Signatures of internal shear layers [Jeong et al., 1997] can be seen, with regions of large strain (which is shear dominated, see section 5.2.2) lying both below vortices, favouring the Q4 side; and above vortices, favouring the Q2 side (see

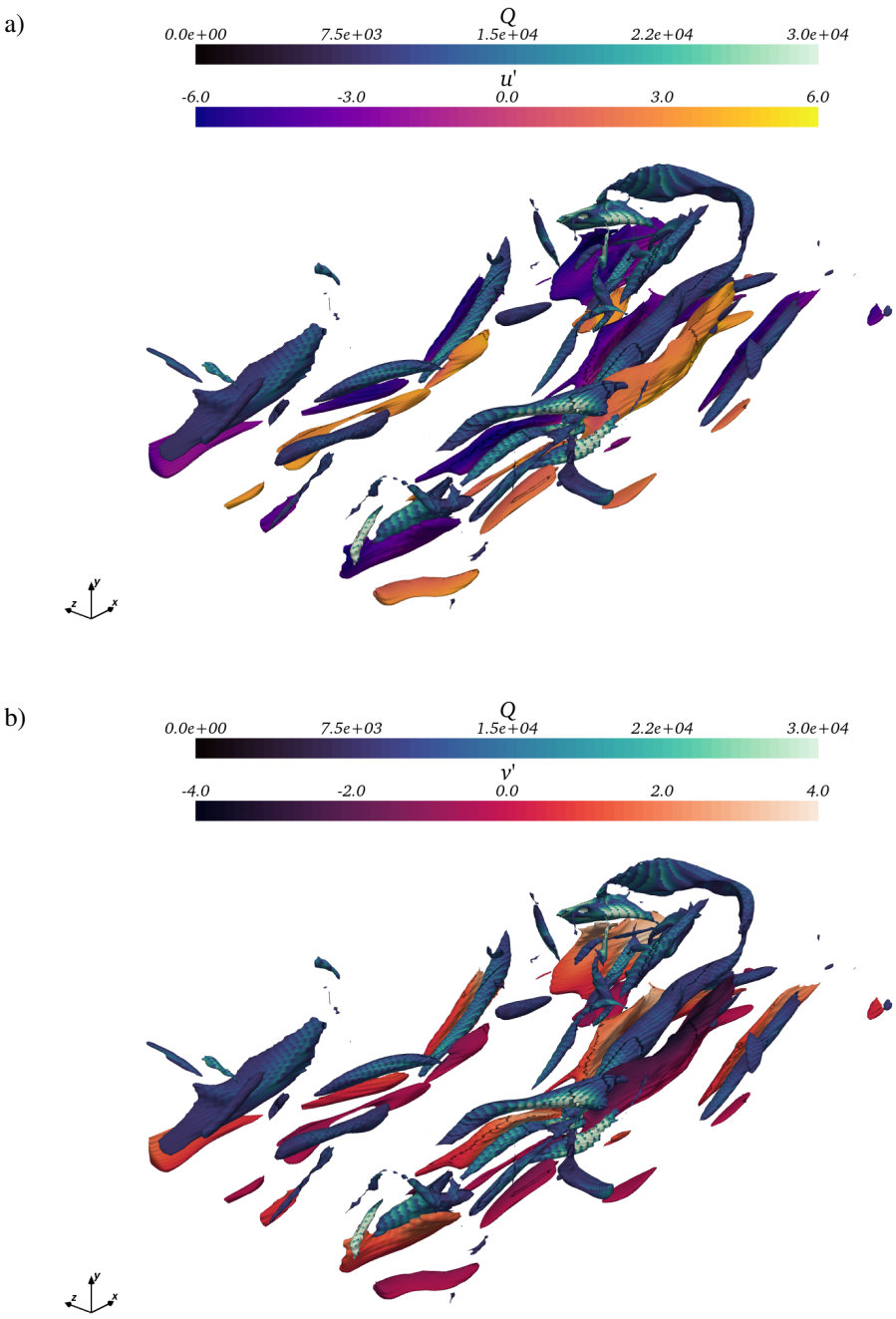


Figure 5.4: Instantaneous 3D snapshot of $C1_\Omega$ and $C1_P$ in the near-wall region, coloured by a) u' and b) v'. $C1_{ns}$ and $C1_S$ are made transparent.

the green enclosures in figure 5.2(b)). Fluid advected below a vortex in a sweep (Q4) is squished, shearing the fast-moving fluid against the wall below. Fluid advected above a vortex in an ejection (Q2) is also squished, as slow-moving fluid hits a "wall" of faster moving fluid above, shearing the ejecting fluid.

$C1_P$ is generally restricted to the the buffer layer and below, where the mean shear is largest. Vortices are most populous in the buffer layer, but extend further into the logarithmic region as well. This can be observed in figure 5.4, where one observes Q2 and Q4 structures flanking quasi-streamwise vortices, and a transverse vortex head extending above the production clusters (see figure 5.3). The mean shear in the production term localizes populations to the near-wall, restricting statistical analysis to the buffer layer and below; therefore, the following section will inspect the rotation-momentum relationship such that structures extend across the wall region.

5.2 Wall-region dynamics

The wall-region ($y^+ < 100$) is examined next, following the same unsupervised learning procedure as in the previous section. This section features additional detailed analyses of the learned clusters, to both validate that the structures learned by the unsupervised learning method are meaningful, and to further establish understanding of wall-region dynamics.

The input data for the wall-region analysis, $\boldsymbol{D}_w$, containing fields of instantaneous $-u'v'$ and Q', progresses through the data pipeline shown prior in figure 4.1, providing $\boldsymbol{M}_w$ containing $k_{best} = 4$ clusters, forming the cluster set $\boldsymbol{C2} = \{C2_1, ..., C2_4\}$. The statistics of each cluster with respect to the input variables are again examined to apply meaningful labels. Cluster $C2_{ns}$ occupies 93% of the channel volume, encompassing what is once-more named the non-salient region of the channel. Two clusters, $C2_{\Omega'}$ and $C2_{S'}$, are characterized by vorticity fluctuation dominated regions ($Q' \gg 0$) and strain fluctuation dominated regions ($Q' \ll 0$), respectively. The fourth cluster, $C2_{-u'v'}$, is characterized by regions of large Reynolds stress, altogether forming the cluster set $\boldsymbol{C2} = \{C2_{ns}, C2_{\Omega'}, C2_{S'}, C2_{-u'v'}\}$. Figure 5.5 shows that u' and v' values within cells labeled as $C2_{-u'v'}$ are distributed neatly within quadrants Q2 and Q4 $C2_{-u'v'}$, which are henceforth denoted by $C2_{Q2}$ and $C2_{Q4}$.

Figure 5.6 demonstrates qualitatively the alignment of clusters with the high intensity regions of their pertinent flow variables. In the near-wall region, the characteristic arrangement of vortices with ejections and sweeps on either side is again observed, with regions of strong fluctuating strain above and below vortices. Q2 and Q4 events in this case extend further above into the outer layer, as the absence of a gradient term in $-u'v'$ leads to $C2_{-u'v'}$ populations being more spread out in the wall-normal direction; though, $-u'v'$ is still generally more intense in the buffer layer as seen in figure 5.5, indicated by the larger distribution spreads.

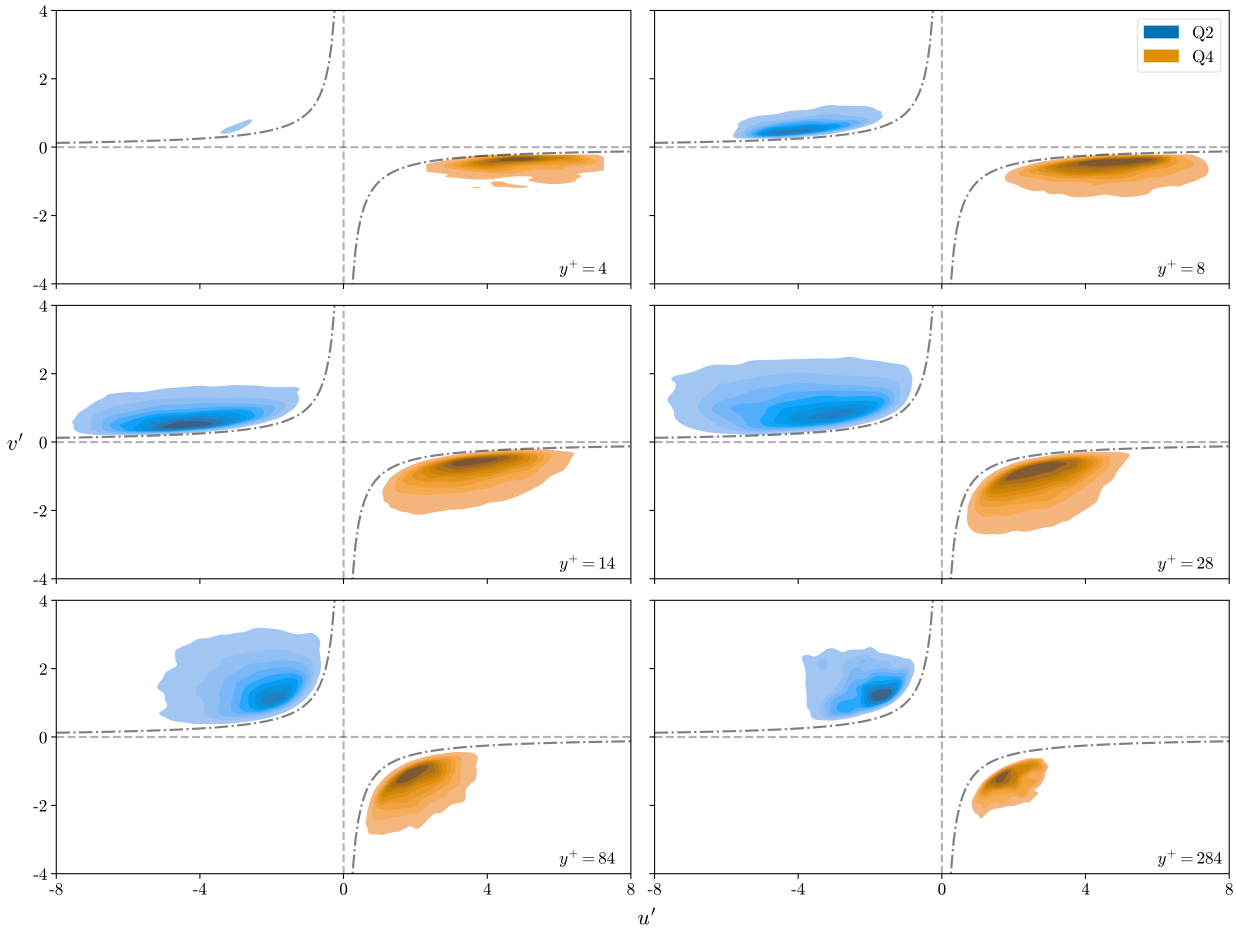


Figure 5.5: Joint $p.d.f(u', v')$ of regions of flow labeled as $C2_{Q2}$ and $C2_{Q4}$ from the a) viscous sublayer, b) lower buffer layer, c) upper buffer layer, d) logarithmic layer, and e) outer layer. Contours contain 90% to 10% of the joint probability mass from the lowest to highest contour levels, respectively. $- \cdot -$: $-u'v' = 1$.

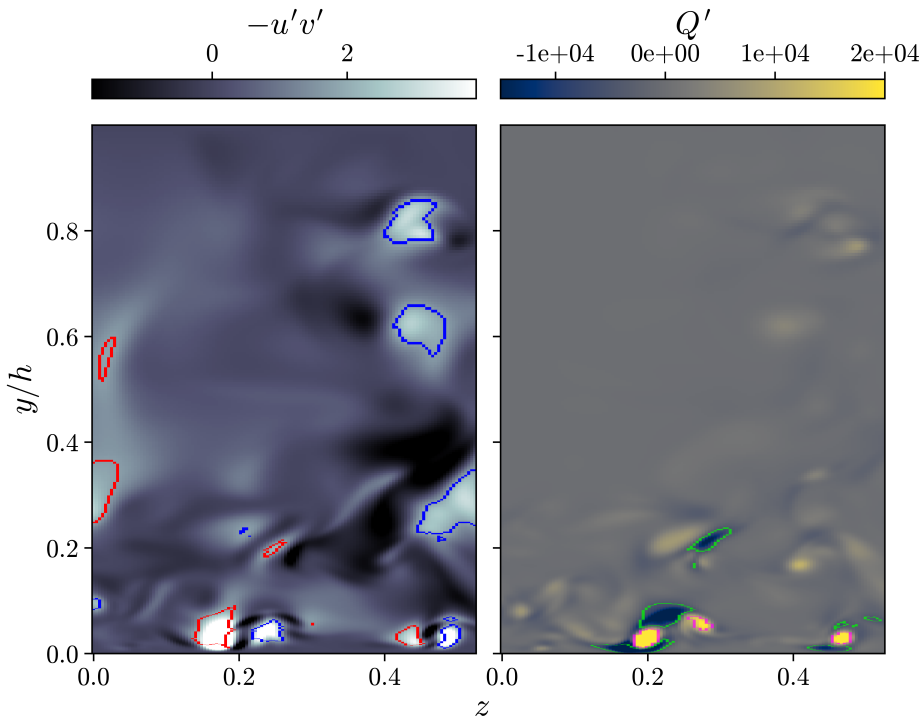


Figure 5.6: Instantaneous cross-stream snapshots of $-u'v'$ (left) and Q' (right) overlaid with cluster boundaries. Red ($C2_{Q4}$), blue ($C2_{Q2}$), pink ($C2_{\Omega'}$) and green ($C2_{S'}$).

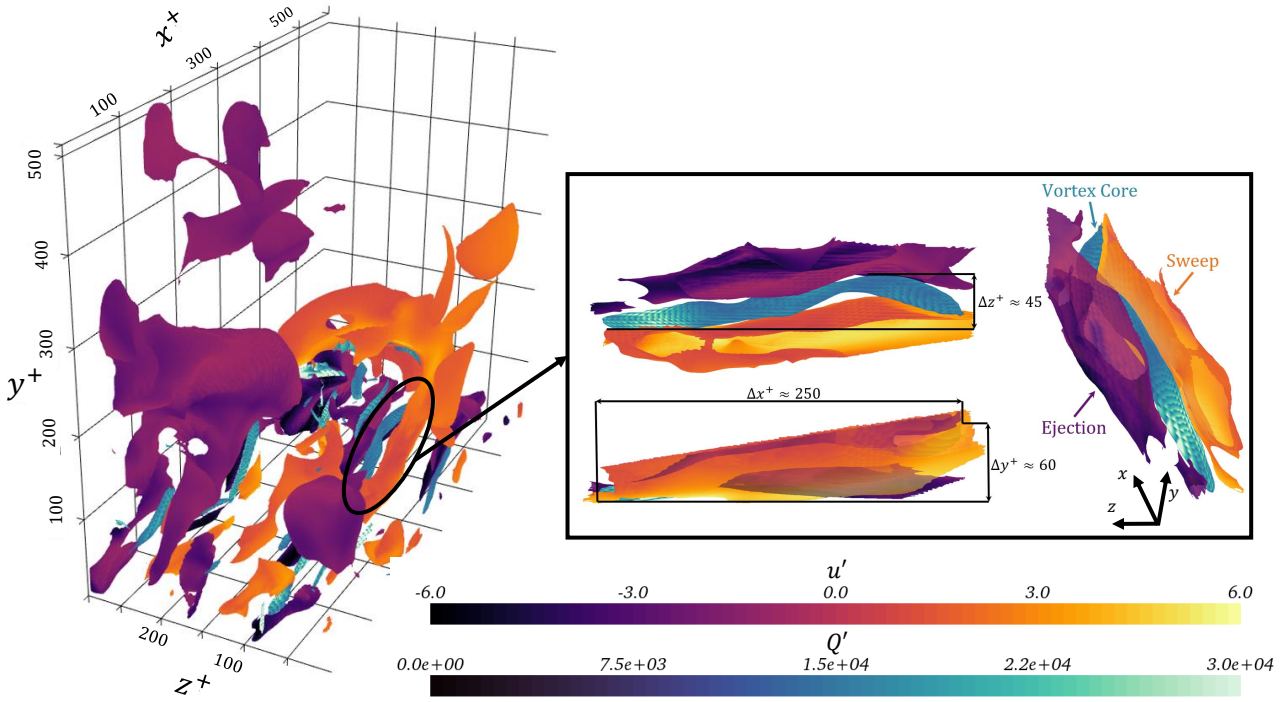


Figure 5.7: Instantaneous sample of momentum transferring structure interactions identified by the unsupervised learning method.

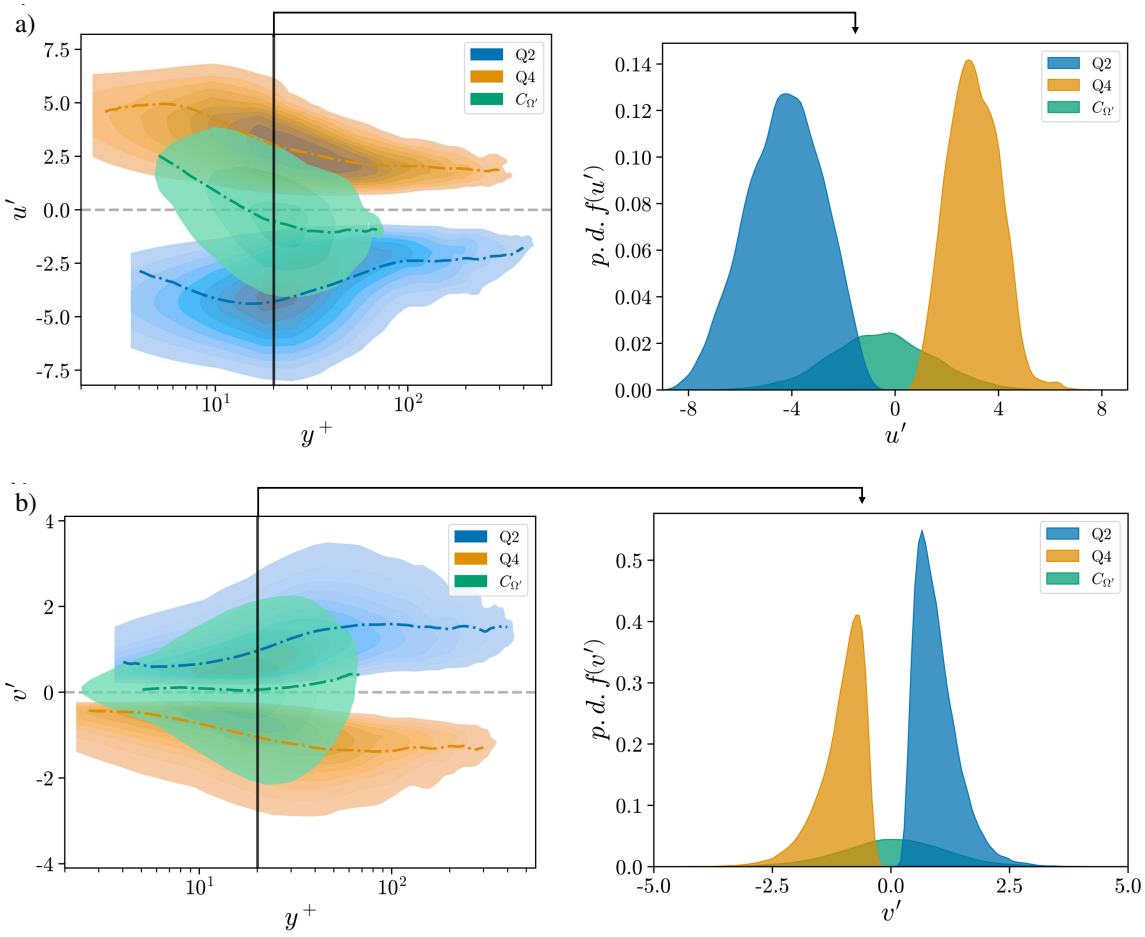


Figure 5.8: Intra-cluster statistics of regions of flow labeled as $C2_{\Omega'}$, $C2_{Q2}$, and $C2_{Q4}$. a) Joint $p.d.f(u', y^+)$ (left) and $p.d.f(u')$ (right) computed at $y^+ \approx 20$. b) Joint $p.d.f(v', y^+)$ (left) and $p.d.f(v')$ (right) computed at $y^+ \approx 20$. Contours contain 90% to 10% of the joint probability mass from the lowest to highest contour levels, respectively. Intra-cluster means are indicated by $- \cdot -$.

Three dimensional visualizations are shown in figure 5.7, with the inset showing a representative vortex, identified by $C2_{\Omega'}$, nestled between an ejection and sweep, identified by $C2_{Q2}$ and $C2_{Q4}$. A sweep occurrence alongside an ejection implies rotation within the flow region cushioned between, in which the three dimensional structures succinctly elucidate this behaviour. The relationship between momentum transferring structures and vortices is represented quantitatively in figure 5.8, showing that the identified vortex populations are generally restricted to the buffer and log layer. Figure 5.8(a) shows that in the early buffer layer ($5 < y^+ < 15$), $C2_{\Omega'}$, i.e., coherent vortices, tend to inherit high u' from the more populated and intense sweeps, whereas in the outer buffer and log layer ($15 < y^+ < 70$), vortices tend to inherit low u' from the more populated and intense ejections; while figure 5.8(b) shows that vortices at all heights generally advect slightly upwards, a signature of vortex lift-up as described by Zhou et al. [1999]. While not all vortices are necessarily situated between an ejection and sweep and streamwise oriented, these statistics seem to suggest that the dominant momentum and energy transferring mechanism in the wall region is that of the vortex flanked by an ejection and sweep. This finding is consistent with the findings of Dong et al. [2017] who found that quadrant events in close proximity to one another tended to form quasi-streamwise trains with a vortex cushioned between. What follows is a further statistical examination of vortex clusters to confirm this claim.

5.2.1 Vortex statistics

Vorticity statistics

The claim that near-wall vortices generally fall between a region of ejecting and sweeping fluid is further explored in this section. The equation for vorticity is given by

$$\boldsymbol{\omega} = \boldsymbol{\nabla} \times \boldsymbol{u}. \tag{5.1}$$

For clarity, the streamwise, vertical, and spanwise components of vorticity in three-dimensional Euclidean space are given by

$$\omega_x = w_{,y} - v_{,z}, \tag{5.2}$$

$$\omega_y = u_{,z} - w_{,x}, \tag{5.3}$$

$$\omega_z = v_{,x} - u_{,y}. \tag{5.4}$$

To first establish an understanding of the physics, a simple schematic of a streamwise vortex flanked by a sweep and ejection will be used, as shown in figure 5.10. Note that, as found earlier in figure 5.5, sweep and ejection strength vary based on wall normal position, but in this simplified model it is assumed that these strengths are comparable; thus, this model can be

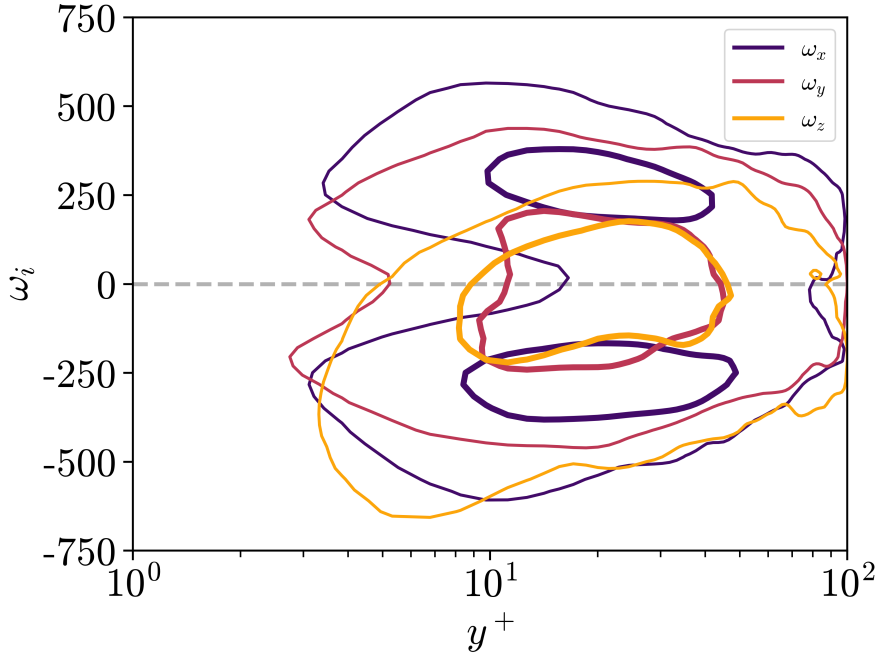


Figure 5.9: Joint $p.d.f(\omega_i^{C2_{\Omega'}}, y^+)$. Contour levels contain 66% (thin line) and 33% (thick line) of the joint probability mass, respectively.

thought of as occurring around the mid-buffer layer, where sweep and ejection intensities are relatively balanced. Furthermore, only a configuration pertaining to a positive streamwise vortex is shown in figure 5.10, where the relative spanwise positions of the ejection and sweep parallel to one another associate with positive streamwise vorticity (ω_x) between; yet, if the ejection and sweep positions were swapped, one would expect negative ω_x to occur between.

The near-wall region is host to both positively and negatively rotating quasi-streamwise vortices [Jeong et al., 1997], as supported by the near-symmetric bi-modal distribution of ω_x within vortex clusters ($C2_{\Omega'}$) shown in figure 5.9. The top contours of the $j.p.d.f$ indicate that vortices are most populous from in the buffer layer. This figure also demonstrates that buffer layer vortices are most characterized by high magnitude ω_x, as indicated by the further extension of the peaks in the ω_x distribution. The vertical vorticity (ω_y) also shows a near-bi-modal distribution (see the outer contour lines of the ω_y $j.p.d.f$), since between ejections and sweeps there is a spanwise gradient of streamwise velocity ($\partial u/\partial z$) that rotates fluid between, whose sign depends on the relative spanwise position of the low-streamwise-speed ejection and the high-streamwise-speed sweep. For the configuration shown in figure 5.10, $\partial u/\partial z > 0$; if the spanwise positions

of the ejection and sweep were swapped the gradient would instead be negative. The spanwise vorticity (ω_z) is near-exclusively negative, due to $\partial u/\partial y$ always being greater than zero due to the mean shear contribution and $\partial v/x$ being negligible in magnitude with respect to $\partial u/\partial y$. This section maintains focus on positively rotating quasi-streamwise vortices; however, results were also obtained for negatively rotating quasi-streamwise vortices and showed similar trends; therefore, those results are omitted to avoid redundancy.

With the simple configuration shown in figure 5.10 established, one can pose the questions: What vortical dynamics are expected to occur between an ejection and sweep? Why do quasi-streamwise vortices persist, despite spanwise vorticity being so prominent near the wall? The instantaneous streamwise vorticity transport equation can be used to investigate these questions

$$\frac{D\omega_x}{Dt} = \omega_j u_{,j} + \cancel{\nabla^2\omega_x} = \omega_x\frac{\partial u}{\partial x} + \omega_y\frac{\partial u}{\partial y} + \omega_z\frac{\partial u}{\partial z} + \cancel{\nabla^2\omega_x} \tag{5.5}$$

where the diffusive term is neglected, as the configuration is assumed to take place in the buffer region where viscous effects are negligible [Brooke and Hanratty, 1993, p. 1017]. Referring to figure 5.10, the expected streamwise vorticity transport term contributions are summarized below:

- $\omega_y\partial u/\partial y$ (tilting):
  - Positive vertical vorticity (ω_y) is generated by way of a positive spanwise gradient of streamwise velocity ($\partial u/\partial z$) between the low speed ejection and high speed sweep
  - ω_y is tilted downwards by $\partial u/\partial y$ into **positive** ω_x
- $\omega_z\partial u/\partial z$ (twisting):
  - Strictly negative ω_z is generated by positive $\partial u/\partial y$
  - ω_z is twisted by positive $\partial u/\partial z$ into **negative** ω_x
- $\omega_x\partial u/\partial x$ (stretching):
  - Positive ω_x is generated due to both negative $\partial v/\partial z$ (and positive $\partial w/\partial y$) between the rising ejection and falling sweep
  - Vortices incline upwards with respect to the wall, thus positive $\partial u/\partial x$ is indirectly adopted, which stretches ω_x into more **positive** ω_x

With these expectations set, one can evaluate the transport statistics of regions identified as vortices to see whether local trends obey theoretical expectations. Cluster $C2_{\Omega'}$ is first filtered

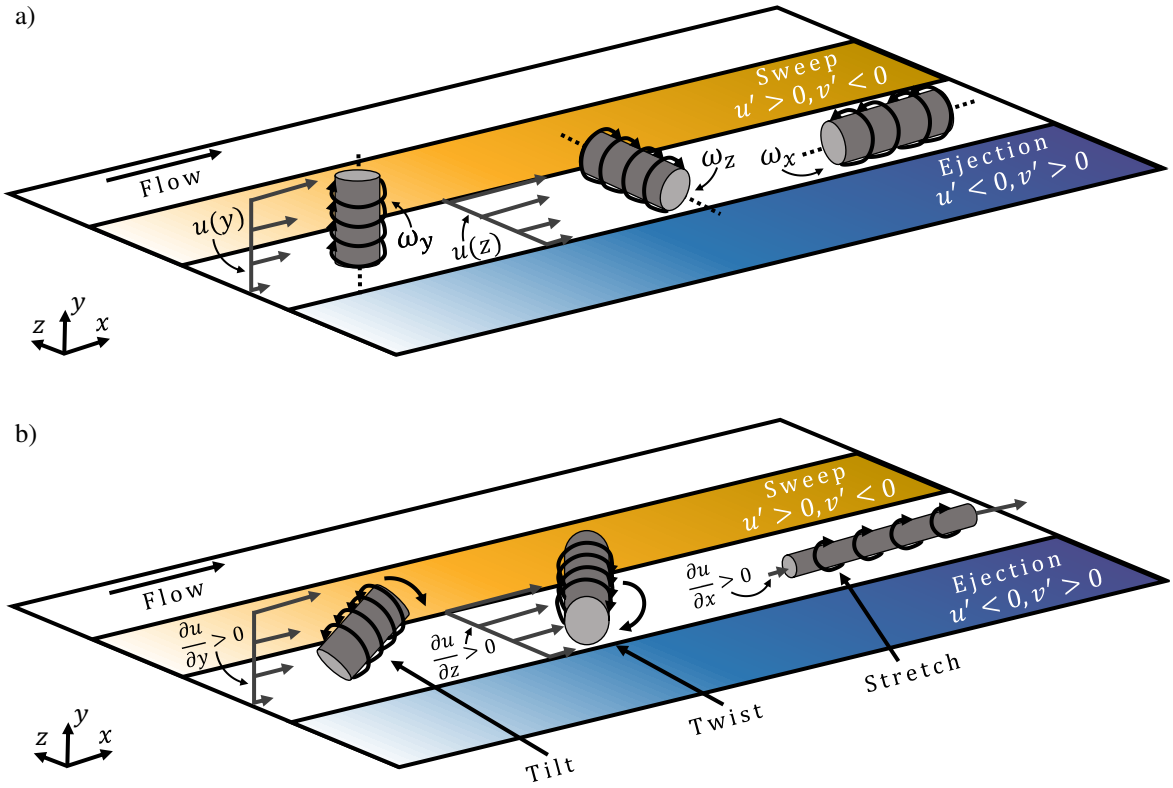


Figure 5.10: Simplified configuration of a near-wall streamwise vortex between an ejection and sweep. The tilting, twisting, and stretching terms of the streamwise vorticity transport are shown at a) some initial position $(\boldsymbol{x}_0, t_0)$, and b) a later position $(\boldsymbol{x}_1, t_1)$ to demonstrate how the velocity gradients within this arrangement influence each transport term.

to consider only positively rotating streamwise vortices ($\omega_x > 0$). Figure 5.11 confirms the previously established streamwise vorticity transport expectations, showing positive contributions to $D\omega_x/Dt$ from vortex stretching and tilting, and negative contributions from vortex twisting. Since the tilting and twisting terms share some components, their net effect can be shown by rewriting equation 5.5 as follows

$$\frac{D\omega_x}{Dt} = \omega_x u_{,x} + \omega_y u_{,y} + \omega_z u_{,z} + \cancel{\nabla^2 \omega_x} \tag{5.6}$$

$$= \omega_x u_{,x} + (u_{,z} - w_{,x}) u_{,y} + (v_{,x} - u_{,y}) u_{,z} \tag{5.7}$$

$$= \omega_x u_{,x} - w_{,x} u_{,y} + v_{,x} u_{,z}, \tag{5.8}$$

where the common $u_{,z}u_{,y}$ terms are cancelled out. The net contributions are shown in figure 5.12, showing a relatively symmetric distribution of twisting, and a generally positive tilting of $-w_{,x}$ by $u_{,y}$. While $u_{,y}$ is generally positive due to the mean shear, $w_{,x}$ is most likely negative since flow between an ejection-sweep region is rotated counter-clockwise (from a top-down perspective in the coordinate system used in this work), such that the leading edge of the vortex tends in the $+z$ direction and the trailing edge tends in the $-z$ direction (as will be shown later in figure 5.14).

The dominant term contributing to the spatio-temporal sustenance and growth of ω_x is that of vortex stretching. An interesting insight is that stretching strength increases nearest to the wall. A possible explanation is that closer to the wall $u_{,y}$ is strongest, and due to the inclined nature of the vortices $u_{,x}$ indirectly adopts the strong wall-normal gradient. As wall-normal position increases, shear strength decreases, leading to a decrease in streamwise vorticity stretching. As will be shown in the following section, vortices tend to advect upwards faster at the downstream tail than at the upstream tail, causing the vortex to lift up and orient more and more vertically near the end of its life cycle, until its eventual decoherence in the logarithmic region and above.

Velocity statistics

Further insights on coherent vortex behaviour is gathered by analyzing how the structures *tend* to advect, by examining the distributions of each fluctuating velocity component within vortex clusters ($C2_{\Omega'}$) over wall-normal position. With figure 5.13 as reference, the following behaviours are demonstrated (note: due to the inclined nature of quasi-streamwise near-wall vortices, it is assumed that the leading edge of a vortex is generally situated closer to the wall, and the trailing edge is generally situated further from the wall):

- Streamwise velocity tends faster at the leading edge and tends slower at the trailing edge of near-wall vortices. This is due to the inheritance of high u' from the more populated and

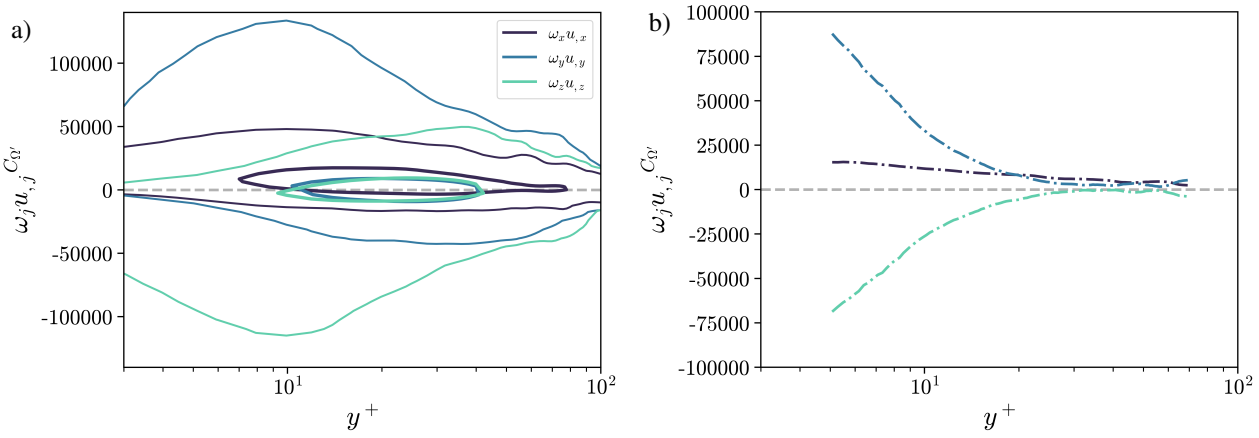


Figure 5.11: a) Joint $p.d.f(\omega_j u_{,j}{}^{C2_{\Omega'}}, y^+)$ with the additional condition of $\omega_x{}^{C2_{\Omega'}} > 0$. Contour levels contain 66% (thin line) and 33% (thick line) of the joint probability mass, respectively. b) Intra-cluster vorticity transport means ($\overline{\omega_j u_{,j}^{C2_{\Omega'}}}$).

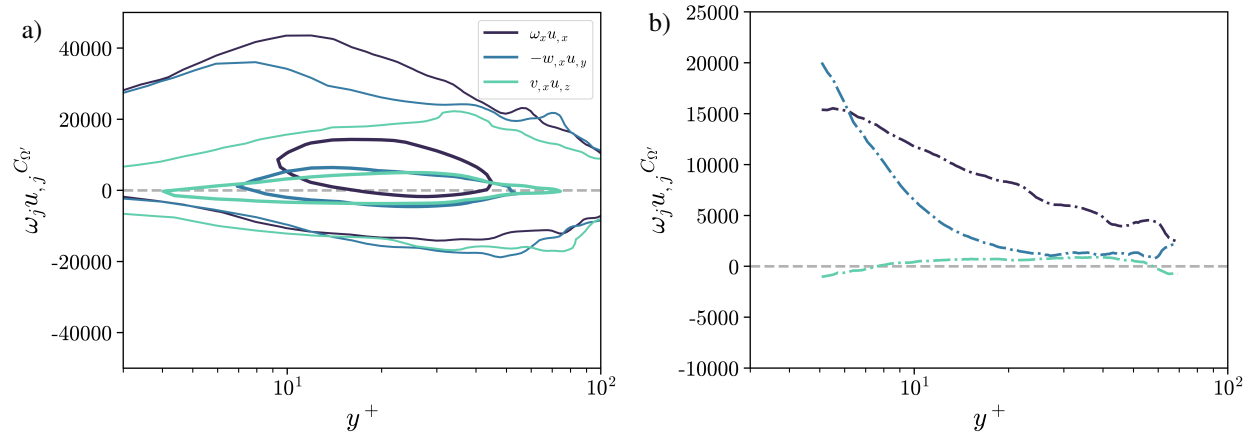


Figure 5.12: a) Joint $p.d.f(\omega_j u_{,j}{}^{C2_{\Omega'}}, y^+)$ using the terms from equation 5.8, with the additional condition of $\omega_x{}^{C2_{\Omega'}} > 0$. Contour levels contain 66% (thin line) and 33% (thick line) of the joint probability mass, respectively. b) Intra-cluster vorticity transport component means.

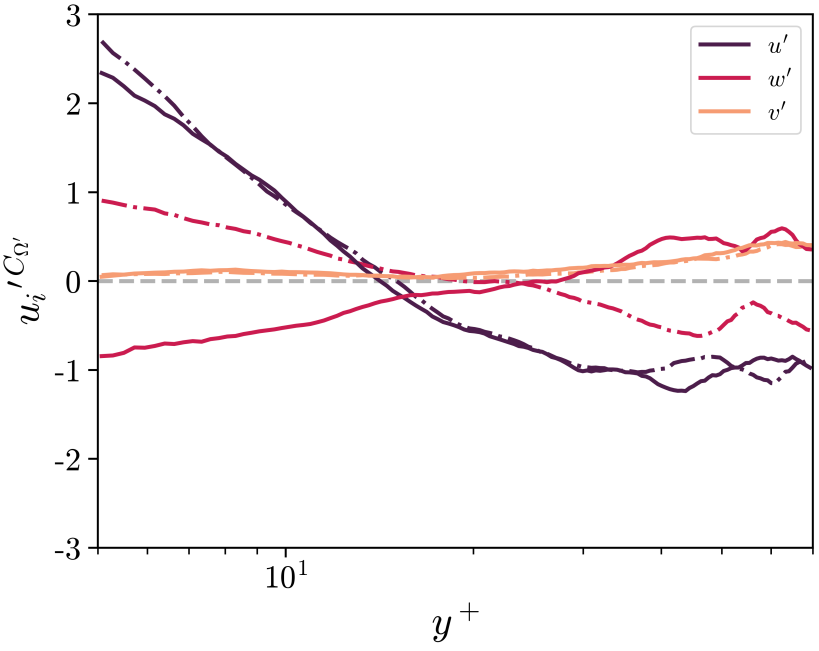


Figure 5.13: Intra-cluster mean velocity fluctuations ($\overline{u_i{}'^{C2_{\Omega'}}}$) over wall-normal height. $-\cdot-$: $\omega_x{}^{C2_{\Omega'}} > 0$; —: $\omega_x{}^{C2_{\Omega'}} < 0$. A qualitative representation of these statistics for a quasi-streamwise vortex with $\omega_x > 0$ is shown in 5.14.

intense sweeps closer to the wall, whereas further from the wall vortices tend to inherit low u' from the more populated and intense ejections

- Wall-normal velocity tends to increase with increasing height. The trailing edge tends upwards more than the leading edge—a signature of the inclination angle commonly observed in coherent vortices [Kim and Moin, 1985]

- Spanwise velocity for coherent vortices, having $\omega_x > 0$, tends in the $+z$ direction at the leading edge, and in the $-z$ direction at the trailing edge. The opposite is found for $\omega_x < 0$. This spanwise velocity imbalance leads to slight spanwise tilts in the vortices (hence *quasi*-streamwise oriented vortices)

A note on the final point is that it is consistent with the findings of Jeong and Hussain [1995, p. 198], who obtained ensemble averaged positively quasi-streamwise oriented rotating vortex (SP) and negatively rotating vortex (SN) structures. The averaged structures were found to overlap, much like a parent-eddy configuration, where the trailing edge of a SP lies on top of the leading

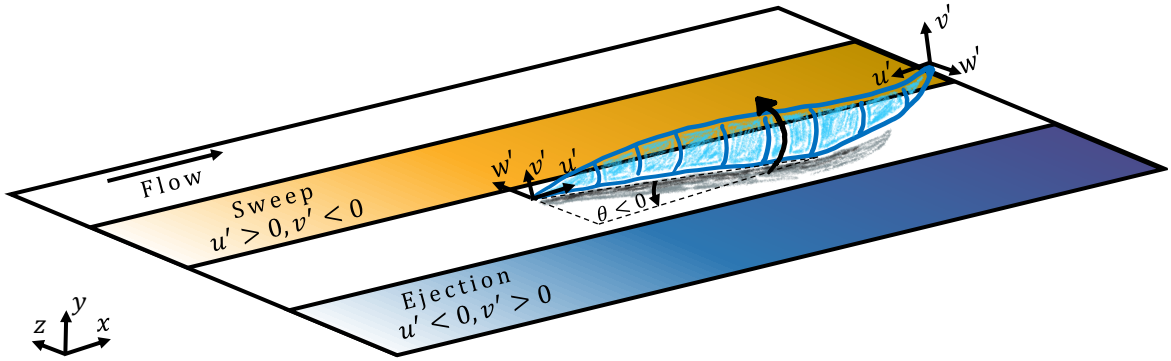


Figure 5.14: Quasi-streamwise positively rotating ($\omega_x > 0$) vortex with velocity vectors shown at its leading and trailing edge.

edge of a SN. Their reasoning for the spanwise velocity gradient along the streamwise extent of the structure was that the counter-rotation of the SP head and the SN tail have an additive effect, such that the counter-rotation thrusts fluid out in the $+z$ direction at the trailing edge of the SP.

While this explanation is plausible given the ensemble averaged structures, there is an alternative explanation. Observing figure 5.14, the leading edge of a positively rotating vortex is rotated in the $+z$ and the trailing edge in the $-z$ by way of negative ω_y that is generated by $\partial u/\partial z > 0$, leading to $\partial w'/\partial x < 0$. This streamwise gradient of spanwise velocity causes a negative spanwise tilt relative to the x-direction. [Jeong et al., 1997, p. 190] find that SP are tilted in the positive direction, and that SN are tilted in the negative direction—note, their positive spanwise direction runs opposite to the one used in this study, so their SP have a negative angle if using the reference frame of this study. This explanation avoids reliance on the overlapping parent-offspring configuration seen in Jeong and Hussain [1995, p. 198], which is plausible for an ensemble averaged CS; however, the specific configuration is less commonly found in complex instantaneous fields. Slight spanwise advection leads to the meandering of vortices, implying the inevitable collision of some of those streamwise propagating vortices with one another, an instrumental part of the self-sustaining process of turbulence [Adrian, 2007, p. 14].

5.2.2 Strain statistics

Regions of substantial strain and negligible rotation, labeled $C2_{S'}$, generally fall below and above coherent vortices (shown qualitatively in figure 5.17). Flow rotating about the vortex is swept downwards into a wall on one side, and ejected into a "wall" of faster moving fluid on the other

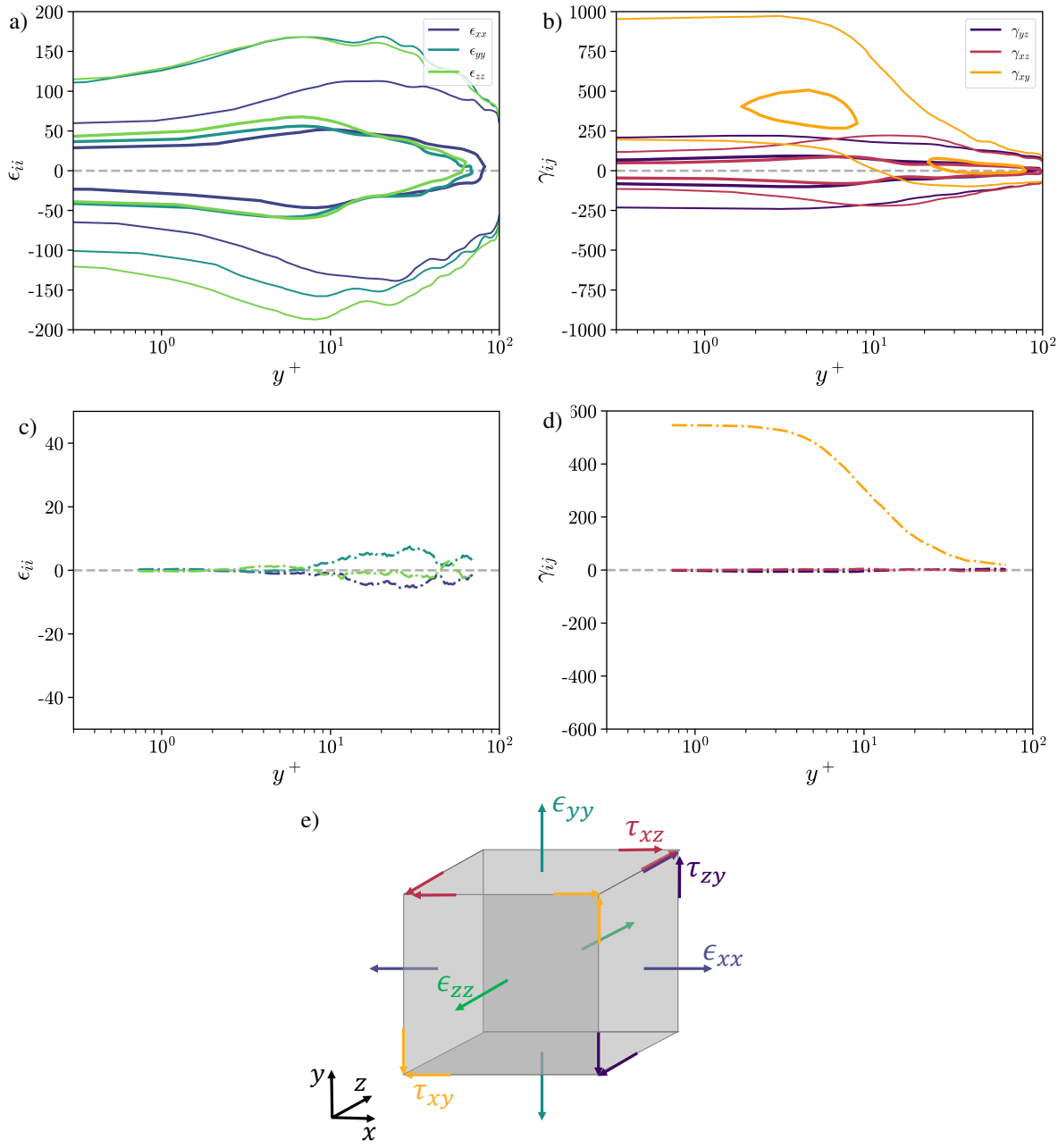


Figure 5.15: Joint $p.d.f(s_{ij}^{C2s'}, y^+)$ for a) normal strain components and b) shear strain components. Contour levels contain 66% (thin line) and 33% (thick line) of the joint probability mass, respectively. Intra-cluster means ($\overline{s_{ij}^{C2s'}}$) are shown for c) normal strain and d) shear strain. Strain schematic on sample grid cell shown in e).

side. Statistics of strain rates within $C2_{S'}$ regions are examined to understand the types of strain experienced by fluid within these regions. The strain rate tensor is given by

$$s_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right) \tag{5.9}$$

or equivalently

$$s_{ij} = \frac{1}{2}\begin{bmatrix} \epsilon_{xx} & \gamma_{yx} & \gamma_{xz} \\ \gamma_{yx} & \epsilon_{yy} & \gamma_{yz} \\ \gamma_{xz} & \gamma_{yz} & \epsilon_{zz} \end{bmatrix} \tag{5.10}$$

where the normal and shear strain components are defined as

$$\begin{aligned} \epsilon_{xx} &= (2u_{,x}), & \tau_{xy} &= (u_{,y} + v_{,x}), \\ \epsilon_{yy} &= (2v_{,y}), & \tau_{xz} &= (u_{,z} + w_{,x}), \\ \epsilon_{zz} &= (2w_{,z}), & \tau_{yz} &= (v_{,z} + w_{,y}). \end{aligned} \tag{5.11}$$

These strain components are shown schematically in figure 5.15(e). Referring to figure 5.15(a-b), fluid straining is generally dominated by τ_{xy}, due to the mean shear. The top contours of the τ_{xy} distribution indicate that regions of intense shear are more populous in the VSL and early-buffer layer ($y^+ < 10$) and the upper-buffer layer ($y^+ > 10$), reinforcing quantitatively the notion that regions of significant shear lie above and below vortices. Other shear components have negligible contributions compared to τ_{xy}. Normal strain distributions are also negligible with respect to τ_{xy}. Though, the larger ϵ_{zz} spread highlights spanwise convergence the i.e., compression, associated with convergences of ejecting fluid; and spanwise dispersion, i.e., tension, associated with inrushes of sweeping fluid [Hwang et al., 2016]; the larger ϵ_{yy} shows a similar wall-normal compressive/tensile nature in regions of high strain as well.

5.3 Inner and outer region dynamics

Regions of intense rotation and associated fluctuating momentous activity are localized to the wall region, yet the dynamics of the outer region certainly influence the dynamics of the inner region [Hwang et al., 2016]. With the flow variables used in the previous two sections, unveiling outer region dynamics is challenging. To bypass the wall-normal inhomogoneity of variables, the variables of interest are standardized by wall-normal dependent root mean square values (equation 3.3). This approach has been used by Del Álamo et al. [2006] and Lozano-Durán et al. [2012] to study vorticity and momentum transferring structures, respectively. Alternative methods to highlight different scales of fluid include low-wavelength filtering in spectral space [Lee et al., 2014] and real space [Motoori and Goto, 2021].

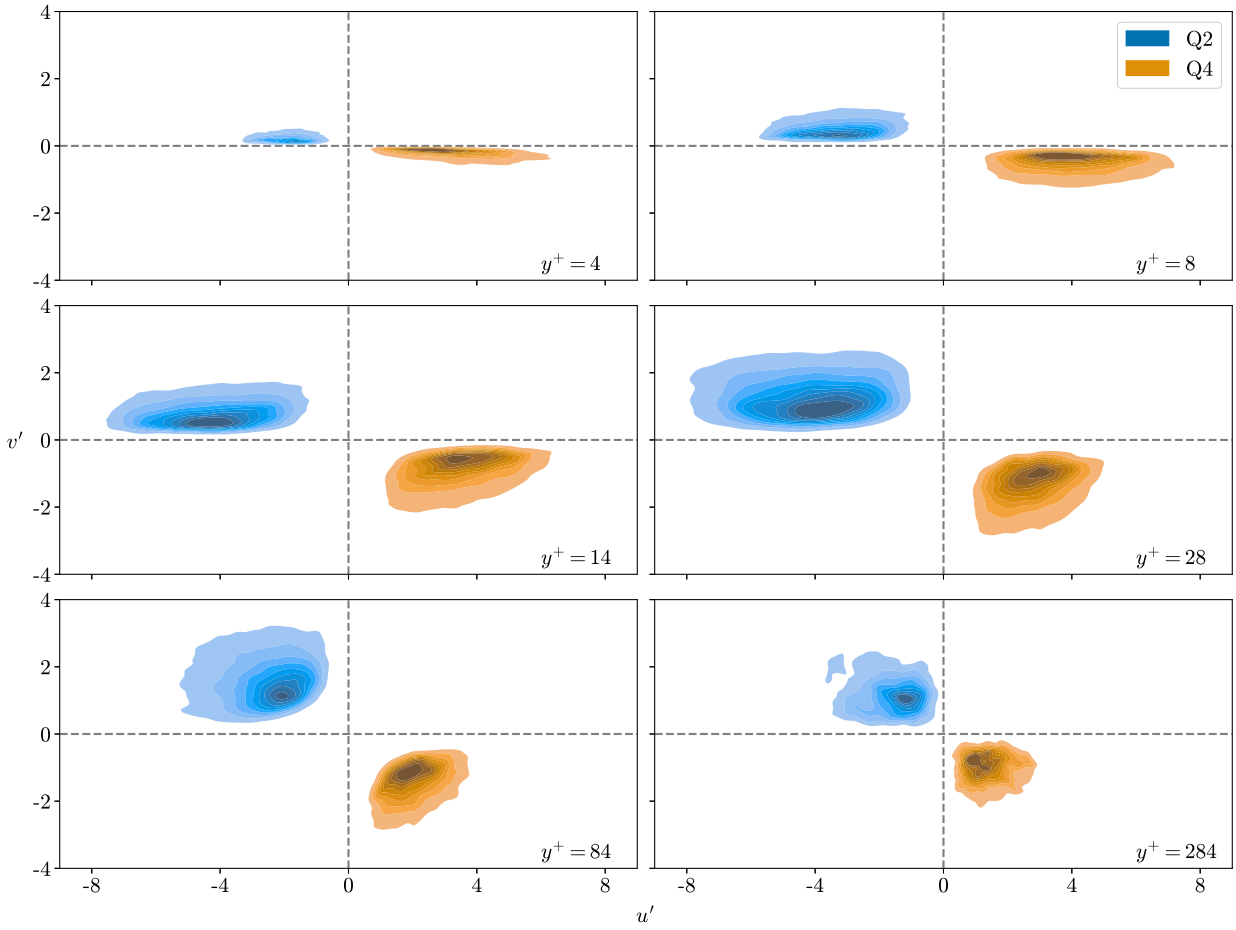


Figure 5.16: Joint $p.d.f(u', v')$ of regions of flow labeled as $C3_{Q2}$ and $C3_{Q4}$ at various channel heights. Contours contain 90% to 10% of the joint probability mass from the lowest to highest contour levels, respectively.

The input data for the full wall-normal region analysis, $\boldsymbol{D}_f$, containing fields of instantaneous $-u'v'^s$ and Q'^s, progresses through the data pipeline shown prior in figure 4.1, providing the mask set $\boldsymbol{M}_f$ containing $k_{best} = 4$ clusters, forming the cluster set $\boldsymbol{C3} = \{C3_1, ..., C3_4\}$. Once again, the statistics of each cluster with respect to the input variables are examined to apply labels. Two clusters, $C3_{\Omega'}$ and $C3_{S'}$, are characterized by vorticity fluctuation dominated regions ($Q'^s \gg 0$) and strain fluctuation dominated regions ($Q'^s \ll 0$), respectively. The cluster, $C3_{-u'v'}$, is characterized by regions of large Reynolds stress. The last cluster, $C3_{ns}$, occupies 74% of the channel volume, encompassing what is once again named the non-salient region of the channel. Altogether, the cluster set $\boldsymbol{C3} = \{C3_{ns}, C3_{\Omega'}, C3_{S'}, C3_{-u'v'}\}$ is formed.

Figure 5.16 shows that velocity values within grid cells labeled by $C3_{-u'v'}$ are distributed within quadrants Q2 and Q4, which are henceforth denoted by $C3_{Q2}$ and $C3_{Q4}$. Near the wall, the sweeps are still generally stronger in the viscous sub-layer and lower buffer layer ($y^+ < 10$), and ejections are still generally stronger in the upper buffer layer and above ($y^+ > 15$). Above the buffer layer the mean shear has a weaker effect on streamwise TKE production, and wall-normal motion is less restricted by the bottom surface; as a result, u' and v' become more balanced, indicated by increasingly circular quadrant distributions as height increases.

Near-wall structures are still maintained after wall-normal standardization, as seen by the similarities in structures when comparing figures 5.6 and 5.17 below $y/h = 0.1$. The outer layer is where standardization brings life to previously unseen structures. In general, structures in the outer layer tend to have larger length scales and time scales (not shown) than those located closer to the wall.

Three dimensional visualizations are shown in figure 5.18, with figure 5.18(a) showing a representative near-wall vortex, identified by $C3_{\Omega'}$, nestled between an ejection and sweep, identified by $C3_{Q2}$ and $C3_{Q4}$. A group of outer layer structures are shown in figure 5.18(b), in which compared to the near-wall structures in figure 5.18(a) the magnitude of u' and Q' are notably lower; length scales are notably larger; and the collection of structures are more structurally isotropic. These points are also demonstrated holistically in figure 5.19, where the instantaneous structures can be seen across the entire spatial domain. In figure 5.19(a), one observes more intense streamwise aligned vortices closer to the wall, and less intense vortices with no apparent preferred orientation in the outer layer.

Figure 5.20(a) shows that below the mid-buffer layer, vortices still tend to inherit high u' from the more populated and intense sweeps, whereas above the mid-buffer layer, vortices tend to inherit low u' from the more populated and intense ejections; while figure 5.20(b) shows that vortices at all heights generally advect slightly upwards, even in the outer layer.

Figure 5.21 confirms that vortices tend to be more streamwise oriented (smaller θ_x) near the wall. In the logarithmic region, θ_z tends to decrease, indicating increasingly spanwise oriented

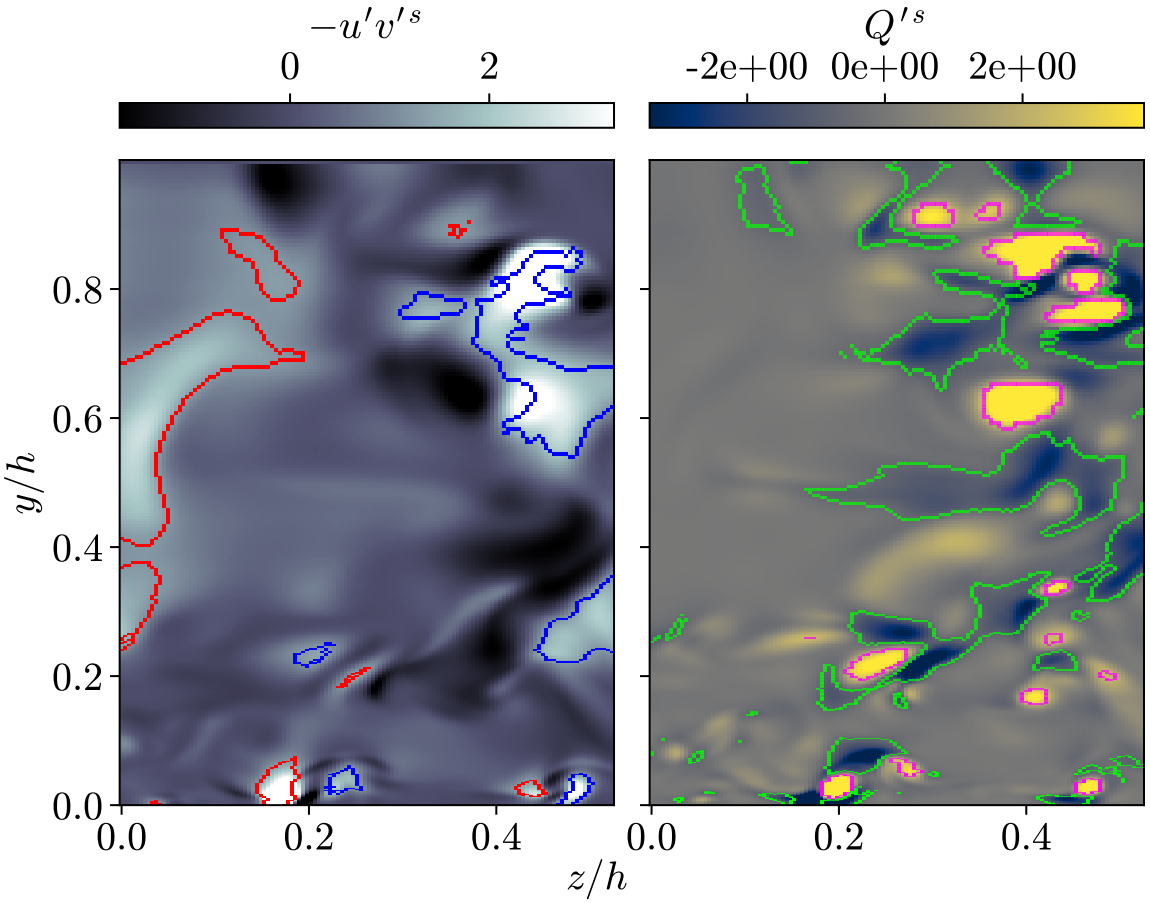


Figure 5.17: Instantaneous cross-stream snapshots of $-u'v'^s$ (left) and Q'^s (right), overlaid with cluster boundaries. Red ($C3_{Q4}$), blue ($C3_{Q2}$), pink ($C3_{\Omega'}$) and green ($C3_{S'}$).

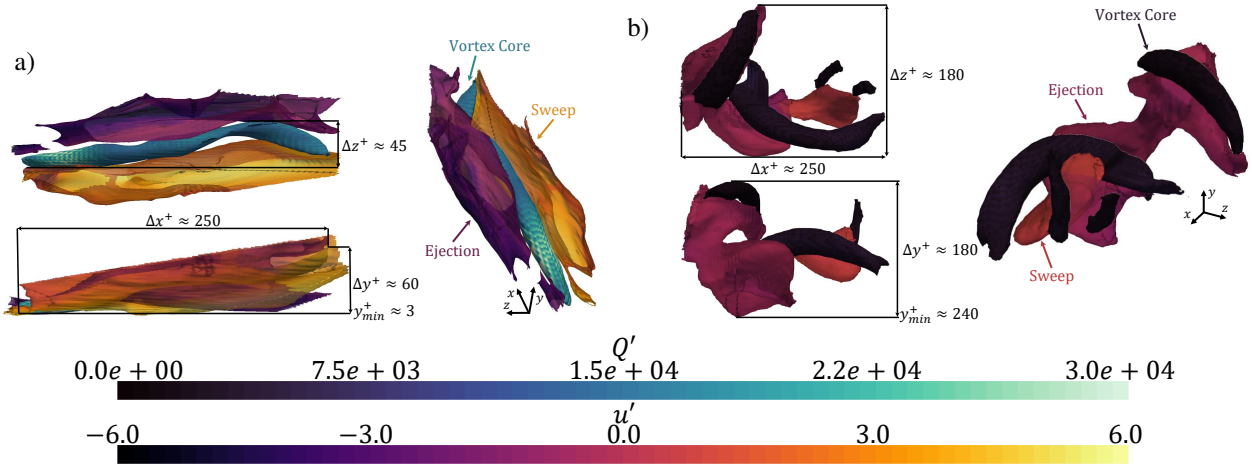


Figure 5.18: Instantaneous top-down, side, and trimetric views of an ejection, sweep, and vortex collection in the a) near-wall region and the b) outer layer.

vortices, highlighting transverse vortices (i.e., the heads of horseshoe vortices commonly observed in this region [Robinson, 1991, p. 625]). As wall-normal distance increases, vortex bounding boxes become more isotropic, consistent with the findings of Motoori and Goto [2021] who found that ensemble-averaged vortices at increasing scales become more spherical. Interestingly, when observing ejection, sweep, and vortex structures occupying the outer layer (figures 5.22–5.28), one can observe flurries of vortical activity between adjacent ejection and sweep regions that occupy significant length scales and persist for significant time scales after their eventual decoherence and dissipation. Figures 5.22–5.28 show a collection of coherent ejecting and sweeping fluid travelling along the periodic channel, beginning with a streamwise extent of $\Delta_x^+ \simeq 500$, eventually stretching out and merging with an upstream ejection (figure 5.24(b) and 5.25(a)) reaching a streamwise extent of $\Delta_x^+ \simeq 1500$, until the eventual decoherence and dissipation of the collection seen in figure 5.28. The behaviour of the low-streamwise-speed regions (purple in the figures) are consistent with the literature of large-scale motions (LSM) and very large-scale motions (VLSM) [Lee et al., 2014, Hwang et al., 2016], e.g., the merging event shown in figure 5.24(b) and 5.25(a) supports the observation that VLSMs tend to be birthed from the merging of multiple LSMs. The aforementioned studies tend to focus on LSMs and VLSMs of slow-moving fluid, this work further illuminates the dynamics by involving regions of fast-moving fluid (sweeps) and intense vortical activity (vortices) in the visualizations as well, showing that a coupled dynamic exists between ejecting LSMs and sweeping LSMs in the outer layer.

Similarities can be seen between these outer layer structures and some of the earlier found near-wall structures, in which the structures enclosed in figures 5.22–5.28 resemble outsized

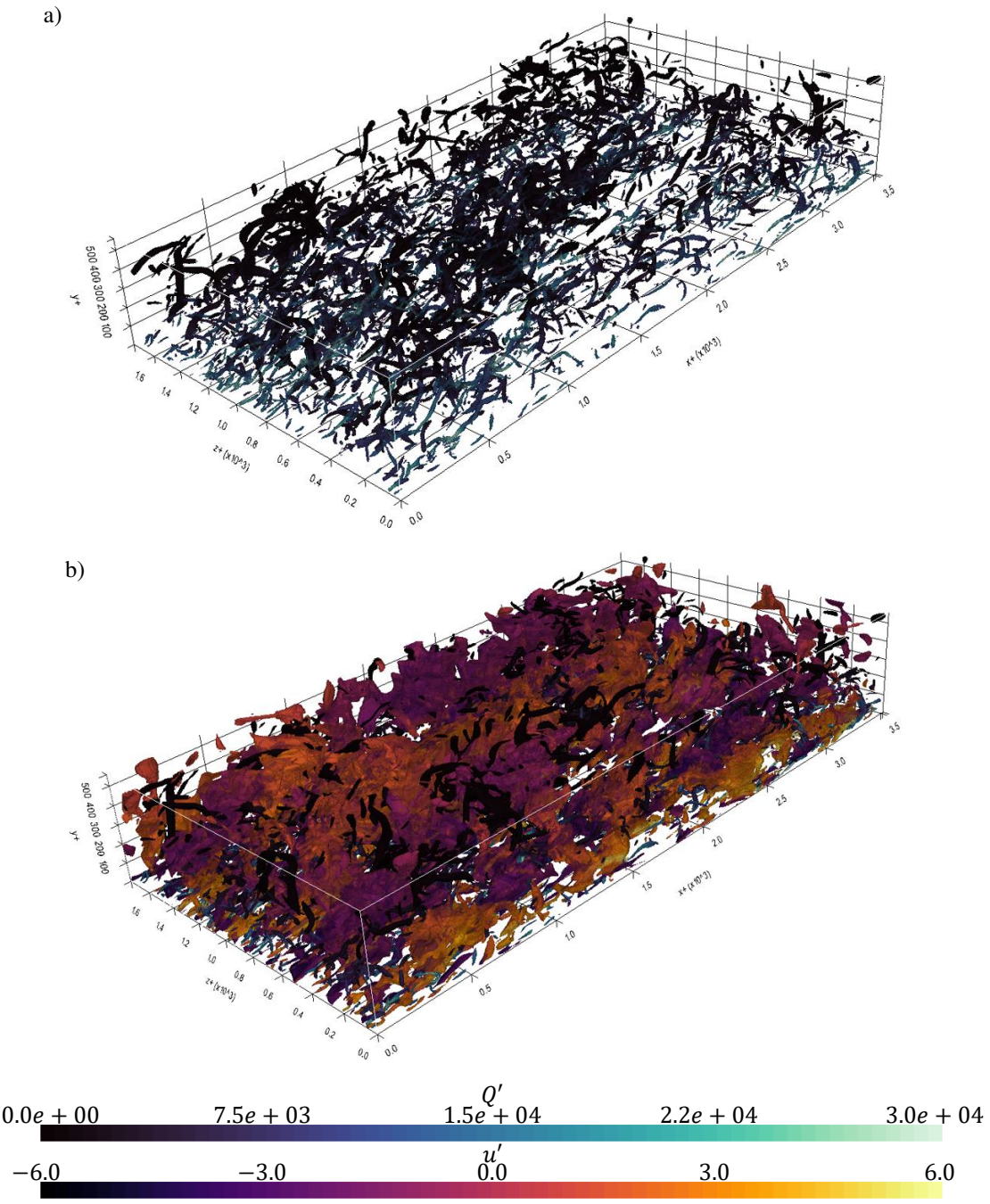


Figure 5.19: Instantaneous a) vortical structures ($C3_{\Omega'}$) and b) vortical and momentum structures ($C3_{\Omega'}$, $C3_{-u'v'}$) across the entire spatial domain of the channel.

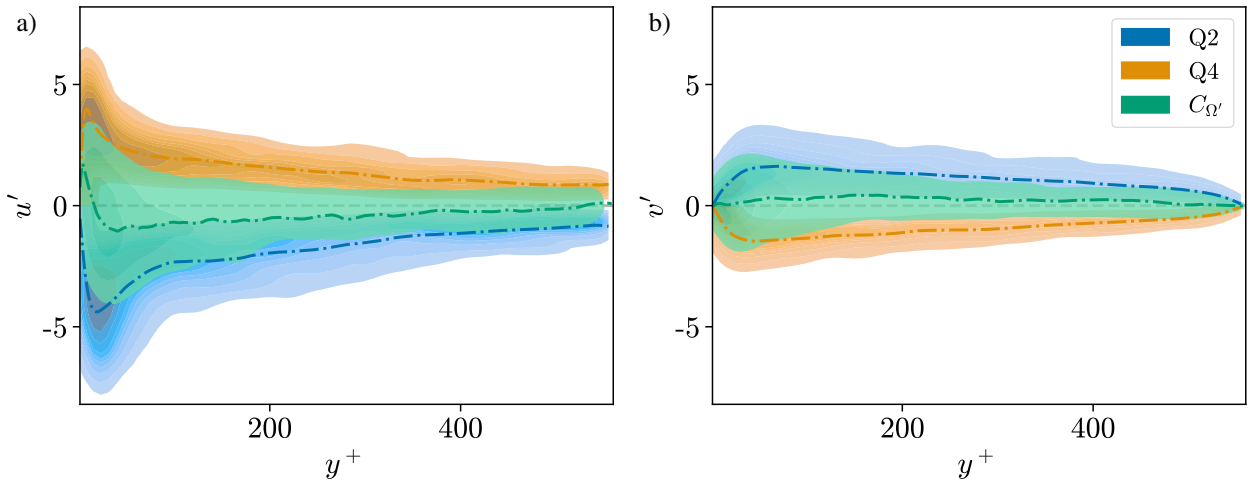


Figure 5.20: Intra-cluster fluctuating velocity statistics of regions of flow labeled as $C3_{\Omega'}$, $C3_{Q2}$, and $C3_{Q4}$. a) Joint $p.d.f(u', y^+)$ and b) joint $p.d.f(v', y^+)$. Contours contain 90% to 10% of the joint probability mass from the lowest to highest contour levels, respectively. Intra-cluster means are indicated by – · –.

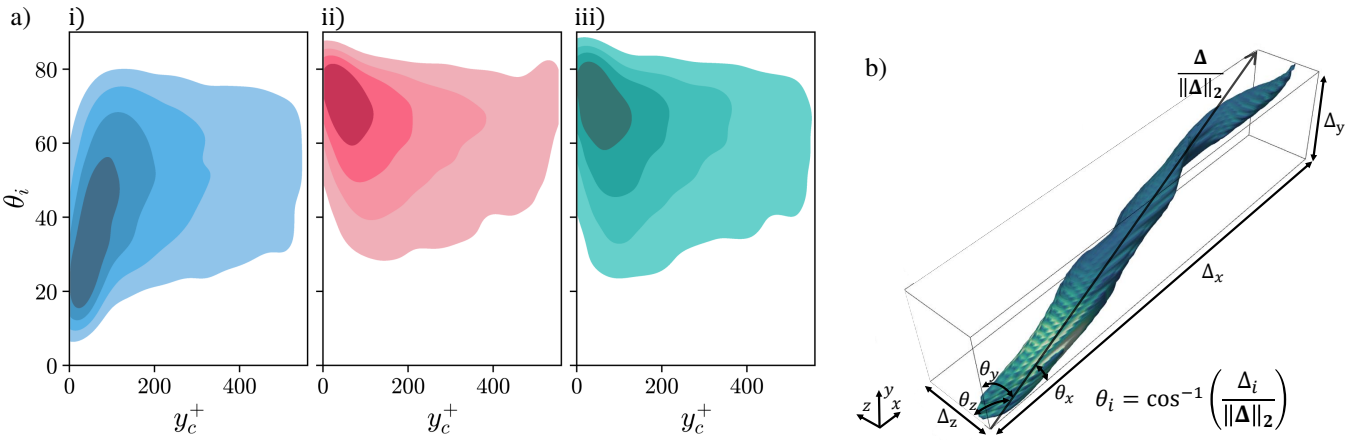


Figure 5.21: a) Intra-cluster bounding box angle statistics of $C3_{\Omega'}$. a) Joint $p.d.f(\theta_i, y_c^+)$ with respect to the i) streamwise, ii) wall-normal, and iii) spanwise directions, where $y_c^+ = y_{min}^+ + \frac{\Delta y^+}{2}$. Contours contain 80% to 20% of the joint probability mass from the lowest to highest contour levels, respectively. b) Sample vortex bounding box schematic.

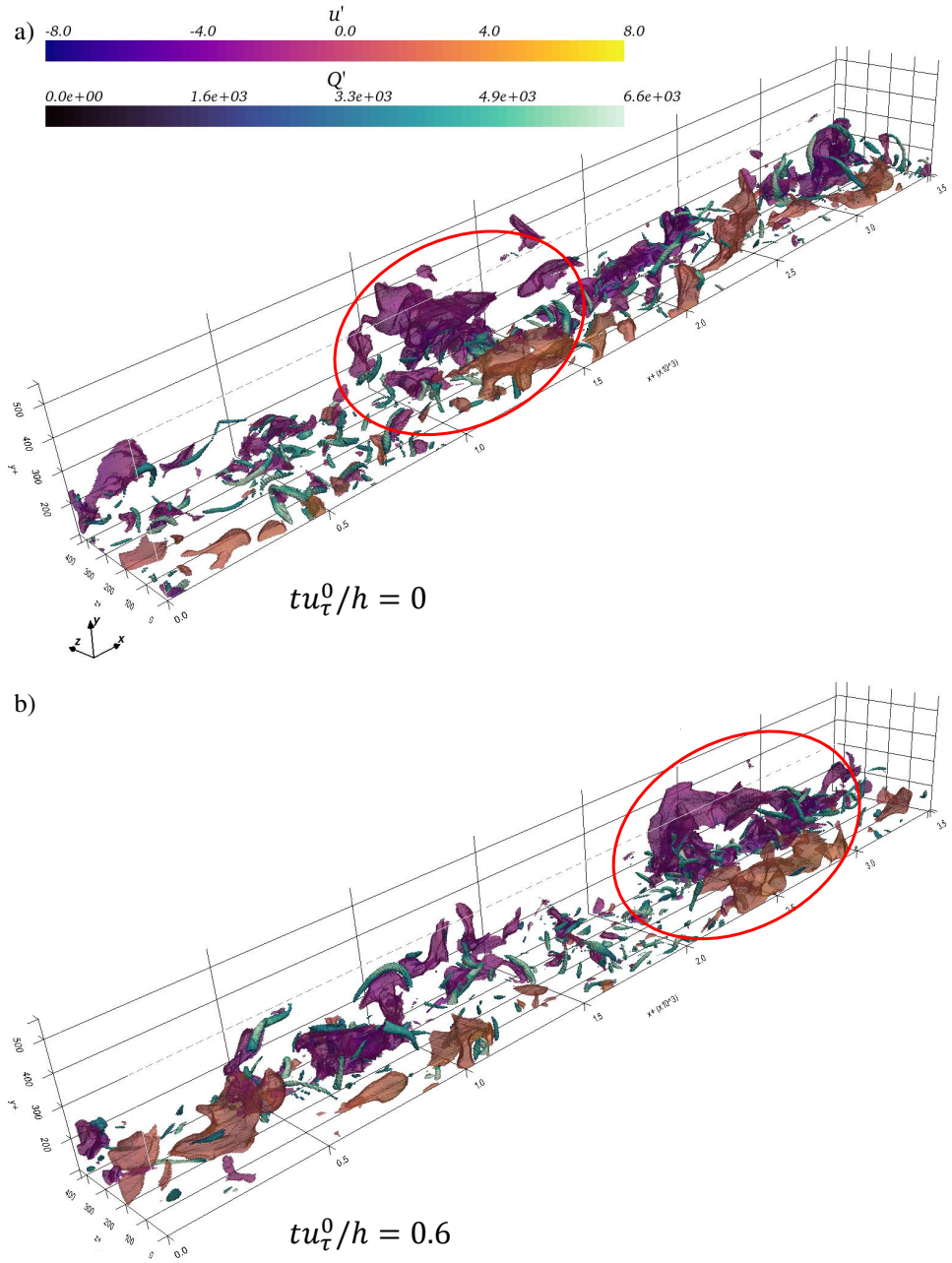


Figure 5.22: Outer layer ($y^+ > 100$) structures at (a) $tu_\tau^0/h = 0$ and (b) $tu_\tau^0/h = 0.6$ across the entire streamwise extent of the channel. Here, $u_\tau^0 = 1$ corresponds to a unit reference velocity. A collection of ejecting (purple) and sweeping (orange) LSMs are enclosed in red, whose streamwise extent is $\Delta_x^+ \simeq 500$.

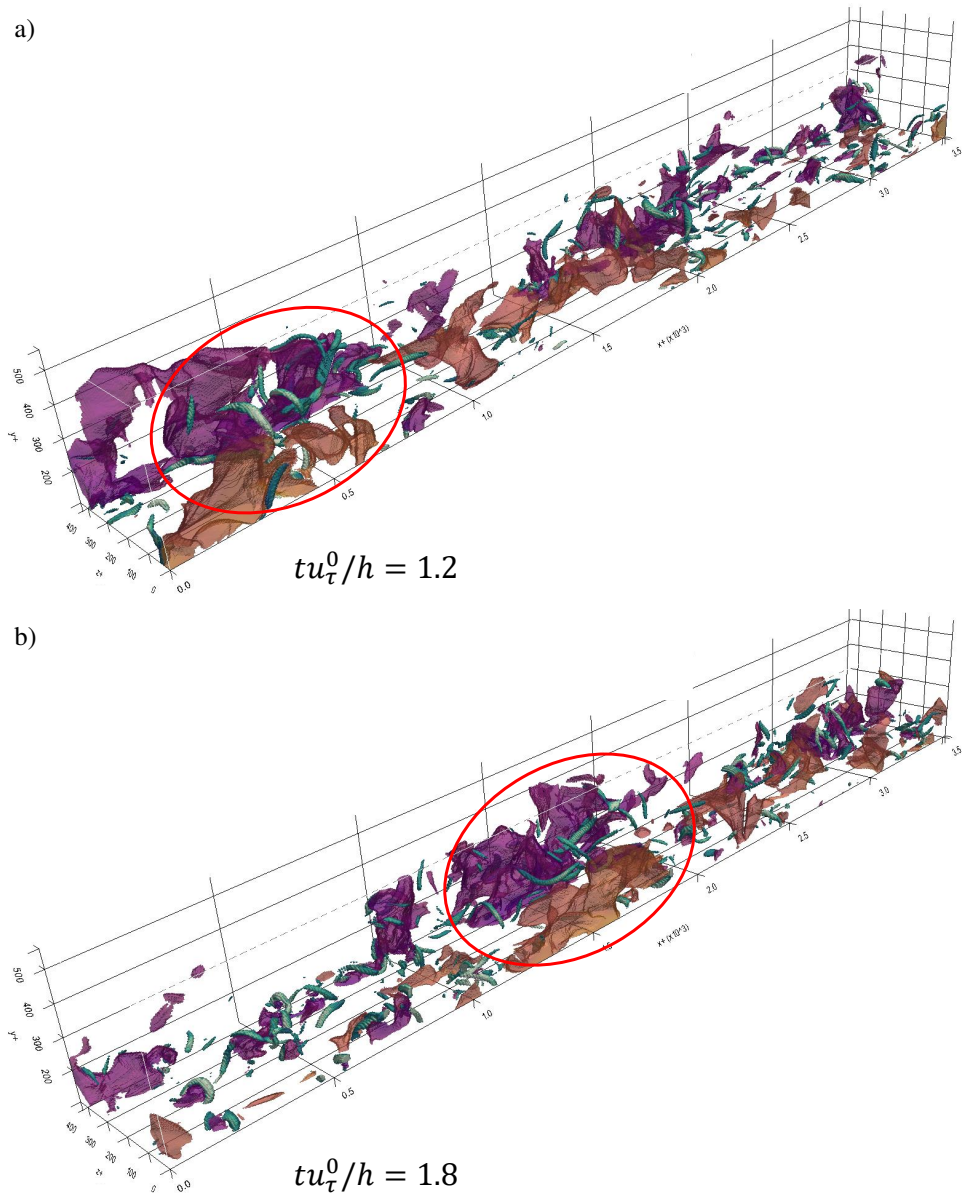


Figure 5.23: Outer layer ($y^+ > 100$) structures (enclosed is the same group of CS shown in figure 5.22) at (a) $tu_\tau^0/h = 1.2$ and (b) $tu_\tau^0/h = 1.8$ across the entire streamwise extent of the channel. The LSMs surround regions of intense vortical activity, and occupy a streamwise extent of $\Delta_x^+ \simeq 750$.

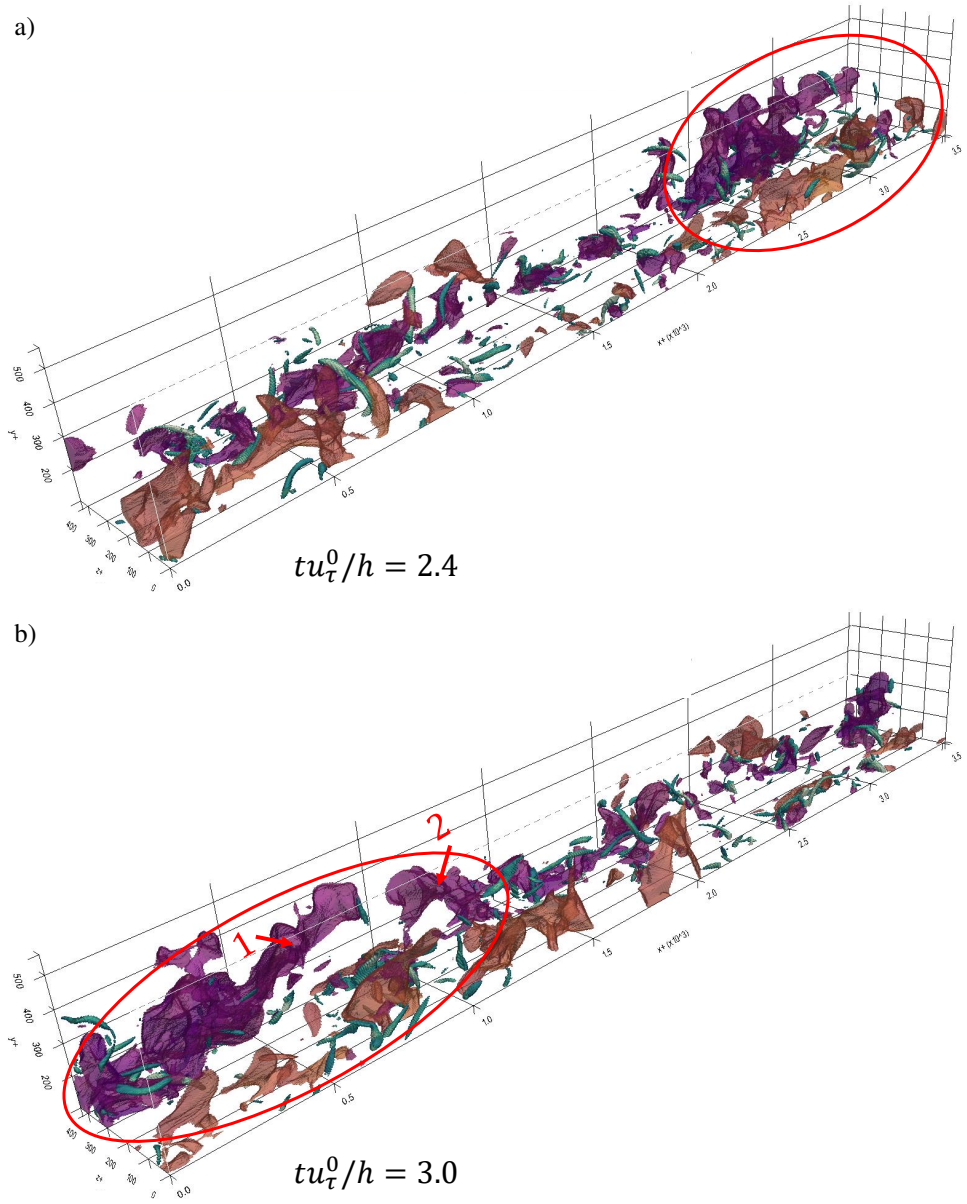


Figure 5.24: Outer layer ($y^+ > 100$) structures at (a) $tu_\tau^0/h = 2.4$ and (b) $tu_\tau^0/h = 3.0$ across the entire streamwise extent of the channel. The LSMs continue to stretch, in which the onset of a merging event between two ejecting LSMs can be seen in b) (labeled 1 & 2).

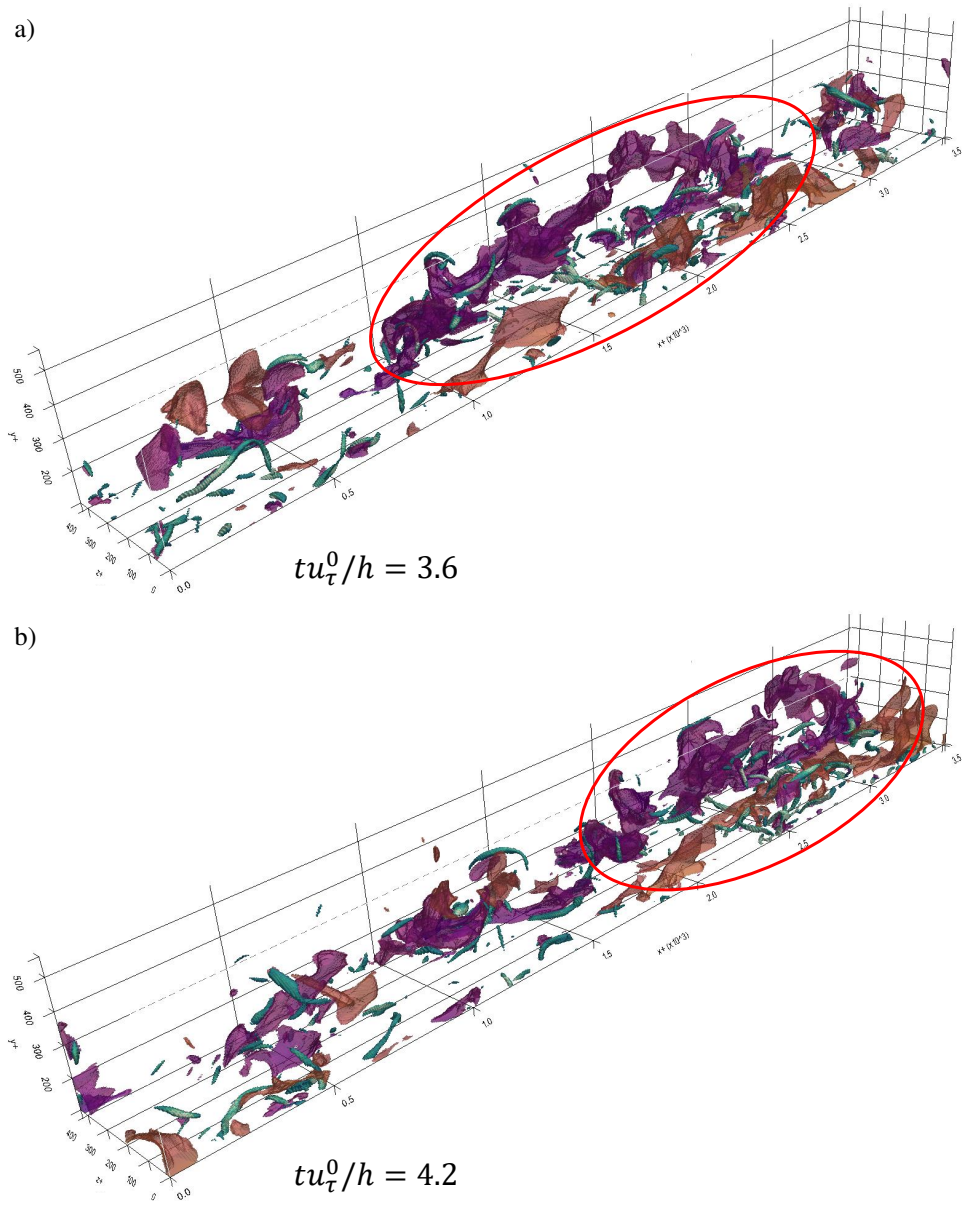


Figure 5.25: Outer layer ($y^+ > 100$) structures at (a) $tu_\tau^0/h = 3.6$ and (b) $tu_\tau^0/h = 4.2$ across the entire streamwise extent of the channel. A VLSM can now be observed, spanning a range of $\Delta_x^+ \simeq 1500$.

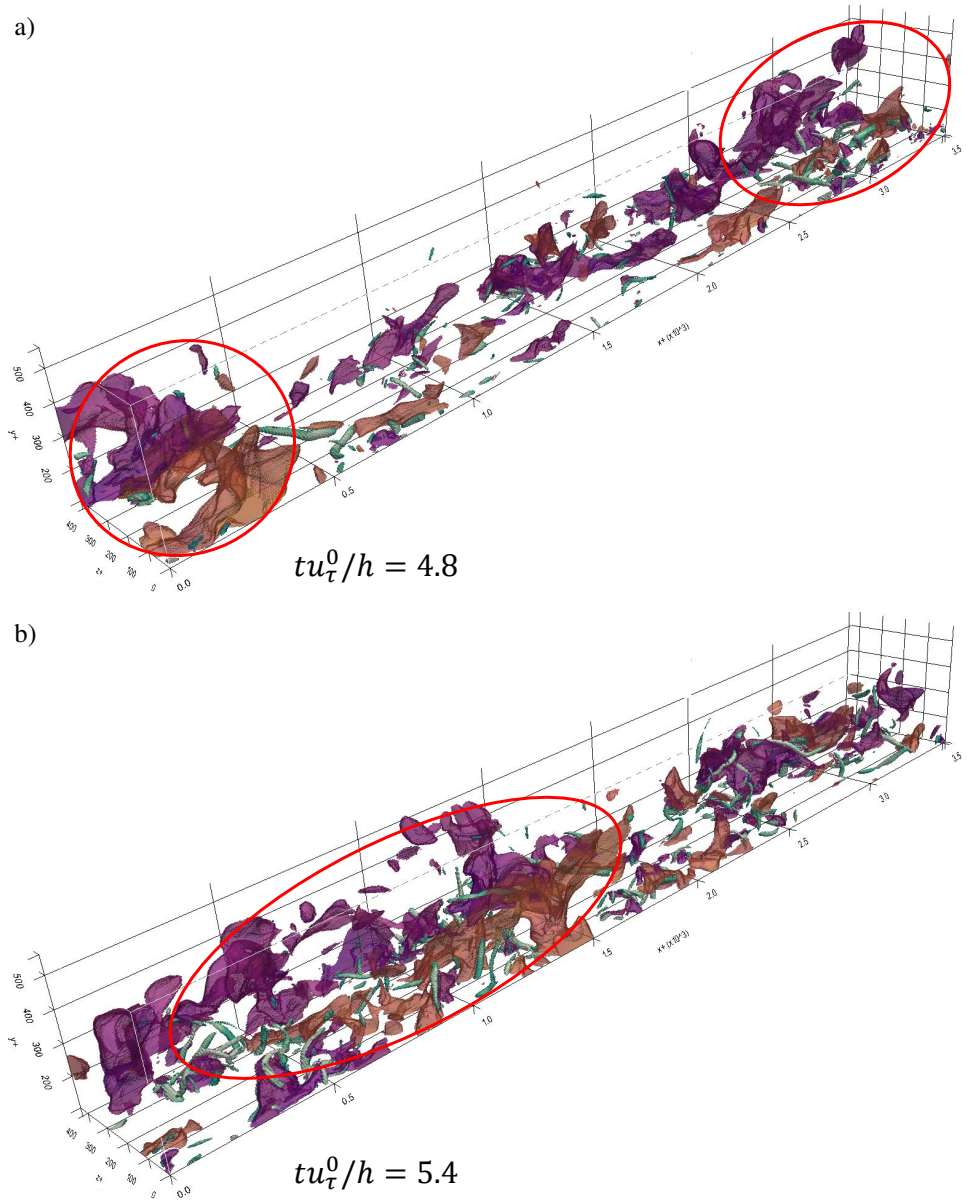


Figure 5.26: Outer layer ($y^+ > 100$) structures at (a) $tu_\tau^0/h = 4.8$ and (b) $tu_\tau^0/h = 5.4$ across the entire streamwise extent of the channel. Hints of decoherence can be observed in b), as the VLSM becomes more disconnected and patchy.

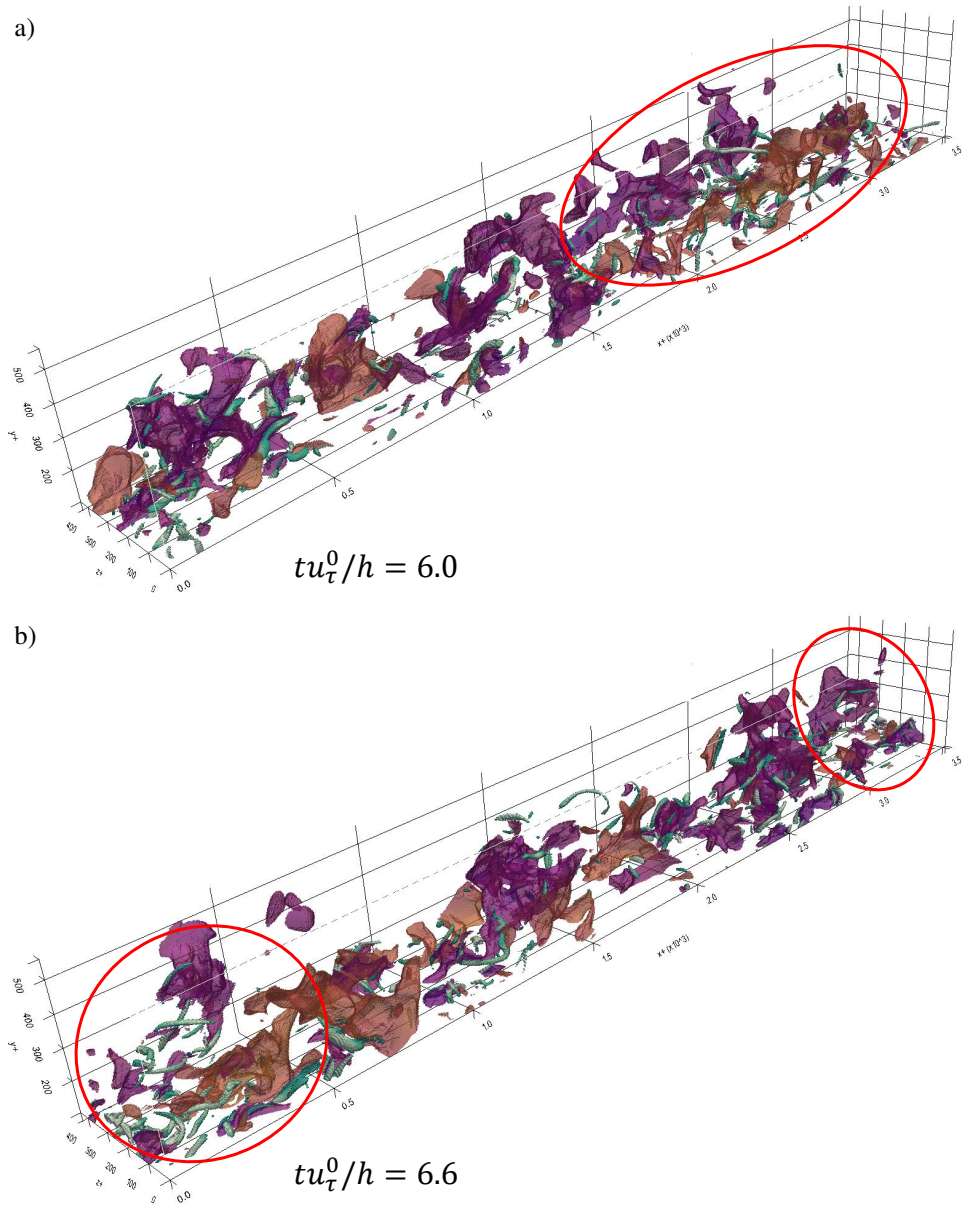


Figure 5.27: Outer layer ($y^+ > 100$) structures at (a) $tu_\tau^0/h = 6.0$ and (b) $tu_\tau^0/h = 6.6$ across the entire streamwise extent of the channel. The decoherence and dissipation of the structures continues.

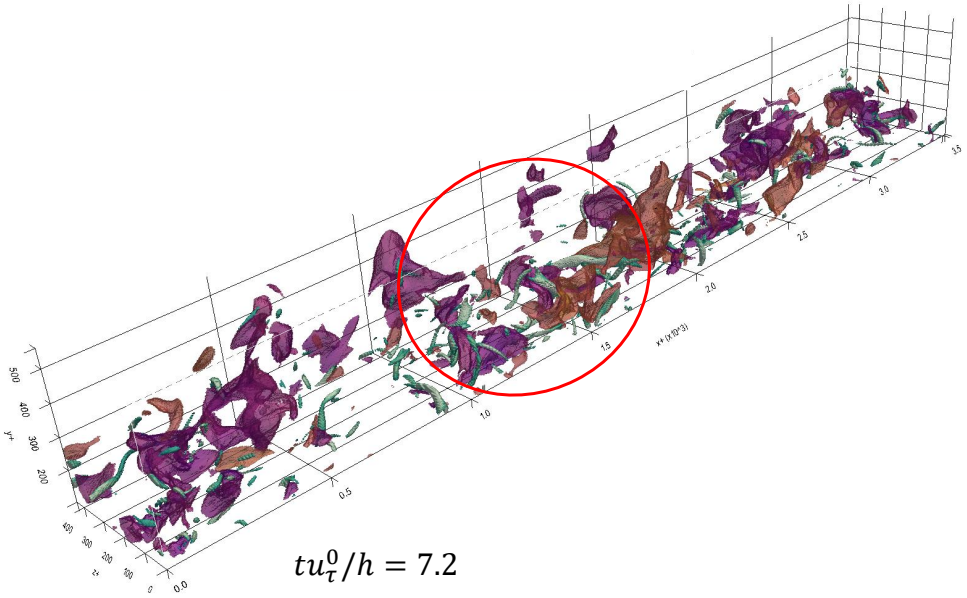


Figure 5.28: Outer layer ($y^+ > 100$) structures at $tu_\tau^0/h = 7.2$ across the entire streamwise extent of the channel. The collection no longer resembles an organized structure, indicating the dissipation of the coherent outer layer LSMs and VLSMs.

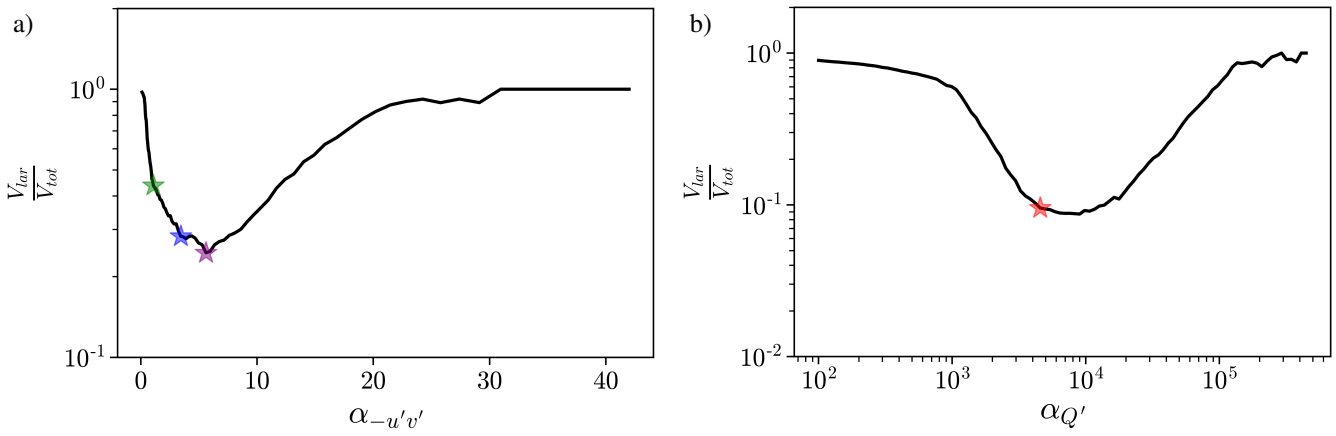


Figure 5.29: Percolation plots for a) Q' fields and b) $-u'v'$ fields. [★, ★, ★] indicate percolation threshold candidates, [$\alpha^1_{perc,-u'v'}$, $\alpha^2_{perc,-u'v'}$, $\alpha^3_{perc,-u'v'}$], for the $-u'v'$ fields. ★ indicates the percolation threshold, $\alpha_{perc,Q'}$, for the Q' fields.

counterparts of near-wall structures (i.e., the inset of figure 5.18), highlighting that there at least exists various scales of the self-sustaining process of wall-bounded turbulence, and that there potentially exists a hierarchy of ejection, sweep, and vortical processes. This may not come to a surprise given what is known about the turbulent energy spectrum, i.e., that "big whirls have little whirls that feed on their velocity, and little whirls have lesser whirls and so on to viscosity" as was elegantly put by [Richardson, 1922, p. 66]. A hierarchy of eddies also implies hierarchies of surrounding momentous fluid, which in channel flow tend to manifest as ejections and sweeps given the streamwise propagation of the flow. This notion coincides with the findings of Motoori and Goto [2021], who investigated hierarchies of low-streamwise-speed regions and vortical regions, finding that lifted low-speed regions (i.e., ejections) flanked by oppositely rotating streamwise vortices exist at various spatial scales within a turbulent channel flow.

5.4 Comparison with percolation theory

In this section, statistics found for the wall-region clusters (section 5.2) identified by the unsupervised method are compared against the statistics of structures identified by percolation theory, followed by qualitative structure comparisons. This section closes with a discussion surrounding the capabilities and limitations of both methods.

5.4.1 Statistical Comparison

First, the percolation process outlined in section 4.4 is applied to the four-dimensional Q' and $-u'v'$ fields used earlier in section 5.2. No pre-processing of the data is required in the percolation process. Structures are found for an array of thresholds for each variable, $Q' > \alpha_{Q'}$ and $-u'v' > \alpha_{-u'v'}$. Logarithmic threshold ranges of $\alpha_{-u'v'} = 0, ..., 45$ and $\alpha_{Q'} = 10^2, ..., 10^6$, each with length 100, were generated, and V_{lar} and V_{tot} are found for the collection of structures identified by each threshold, yielding figure 5.29. Figure 5.29(a) demonstrates a limitation of the percolation method, providing a case in which the percolation threshold is ambiguous. Three potential percolation threshold candidates are shown, $[\alpha^1_{perc,-u'v'}, \alpha^2_{perc,-u'v'}, \alpha^3_{perc,-u'v'}] \simeq [1.1, 3.9, 5.6]$, that come either after a steep slope, or after a steep slope and within a local minimum.

The buffer layer quadrant distributions from figure 5.5 are compared against the quadrant distributions of the structures found using the three thresholds of figure 5.29(a), this time only showing two contour levels to ease comparison. Figure 5.30 shows that the three quadrant distributions of the percolation structures surround that of the unsupervised method, with regions identified with $\alpha^1_{perc,-u'v'}$ having distributions that hug closer to the quadrant origin (figure 5.30(a)), indicating less intense Reynolds stress; whereas regions identified with $\alpha^2_{perc,-u'v'}$ and $\alpha^3_{perc,-u'v'}$ are distributed further from the quadrant axes (figure 5.30(b–c)), indicating regions with more intense Reynolds stress when compared to the decision boundary found by the unsupervised method. Nevertheless, the general trend of early buffer layer sweep strength and late buffer layer ejection strength is found in both methods. Based on figure 5.30, the decision threshold of the unsupervised method lies between $\alpha^1_{perc,-u'v'}$ and $\alpha^2_{perc,-u'v'}$.

For the fields of Q', the percolation threshold is used to identify regions where $Q' > \alpha_{perc,Q'}$, where $\alpha_{perc,Q'} \gg 0$, i.e., regions where rotation dominates strain. The vorticity component distributions shown in figure 5.9 are compared with the distributions returned by the percolation structures. Figure 5.31 shows that the distributions are similar, with the vorticity distributions belonging to the unsupervised method extending slightly further out. The populations of percolation structures generally have smaller ω_x and slightly smaller ω_y, yet the shapes of the distributions are consistent, signalling that the general statistical trends found earlier (i.e., a bi-modal distribution of streamwise vorticity in the buffer layer) comply when using either method. Note, the greater reduction in streamwise vorticity indicates that regions of large Q', i.e., quasi-streamwise vortices, are most characterized by ω_x. These distributions suggest that the vortices identified by the unsupervised method highlight vortices closer to their high intensity cores, hence the more intense vorticity components; whereas percolation theory returns structures, identifying vortices slightly further from their cores (as is shown qualitatively in figure 5.32). Based on figure 5.31, the decision threshold of the unsupervised method likely lies further down the plateau of the

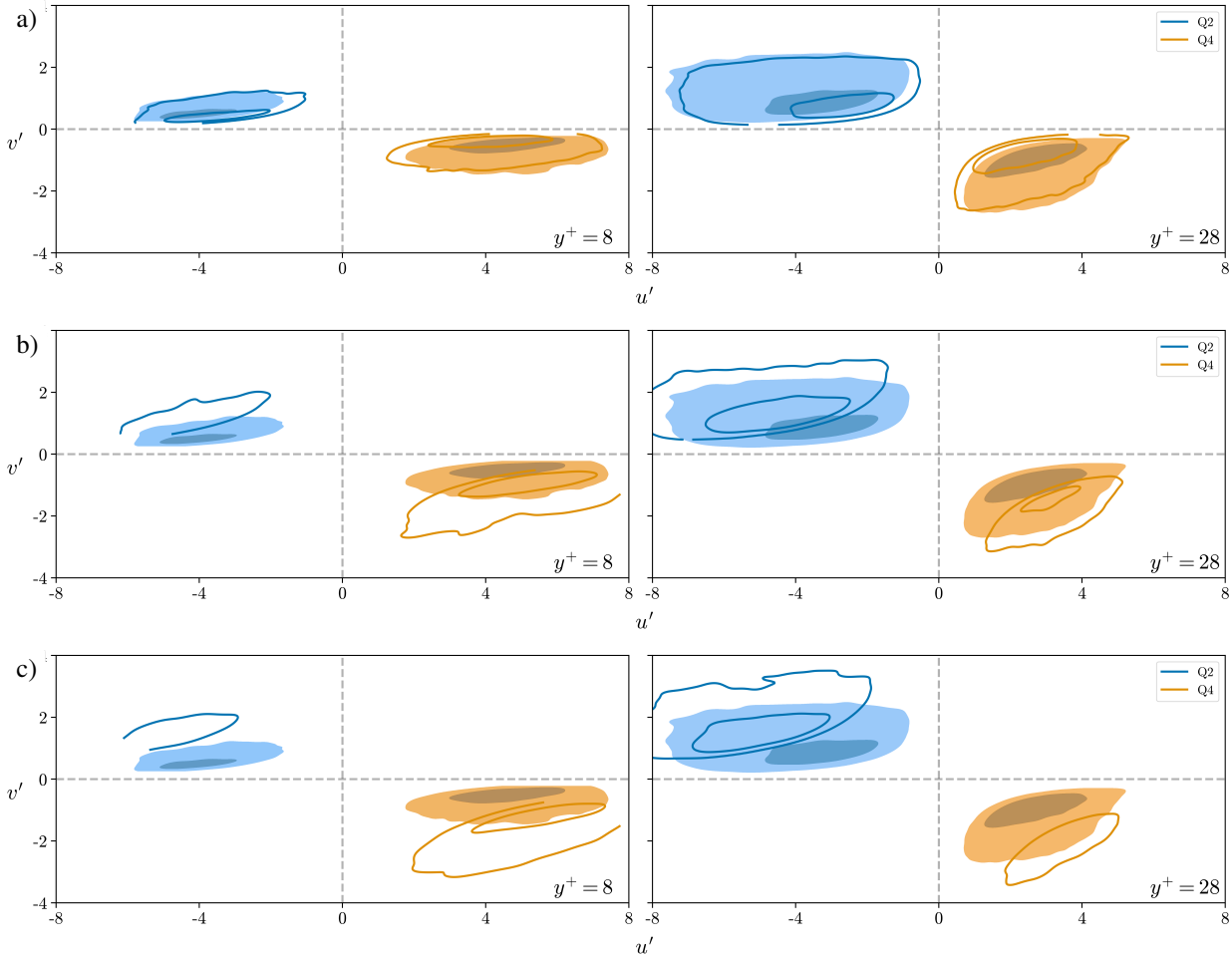


Figure 5.30: Joint $p.d.f\,(u', v')$ of clusters $C2_{Q2}$ and $C2_{Q4}$ (filled) and structures identified using a) $\alpha^1_{perc,-u'v'} = 1.1$, b) $\alpha^2_{perc,-u'v'} = 3.9$, and c) $\alpha^3_{perc,-u'v'} = 5.6$ (lines). Contour levels contain 66% and 33% of the joint probability mass, respectively.

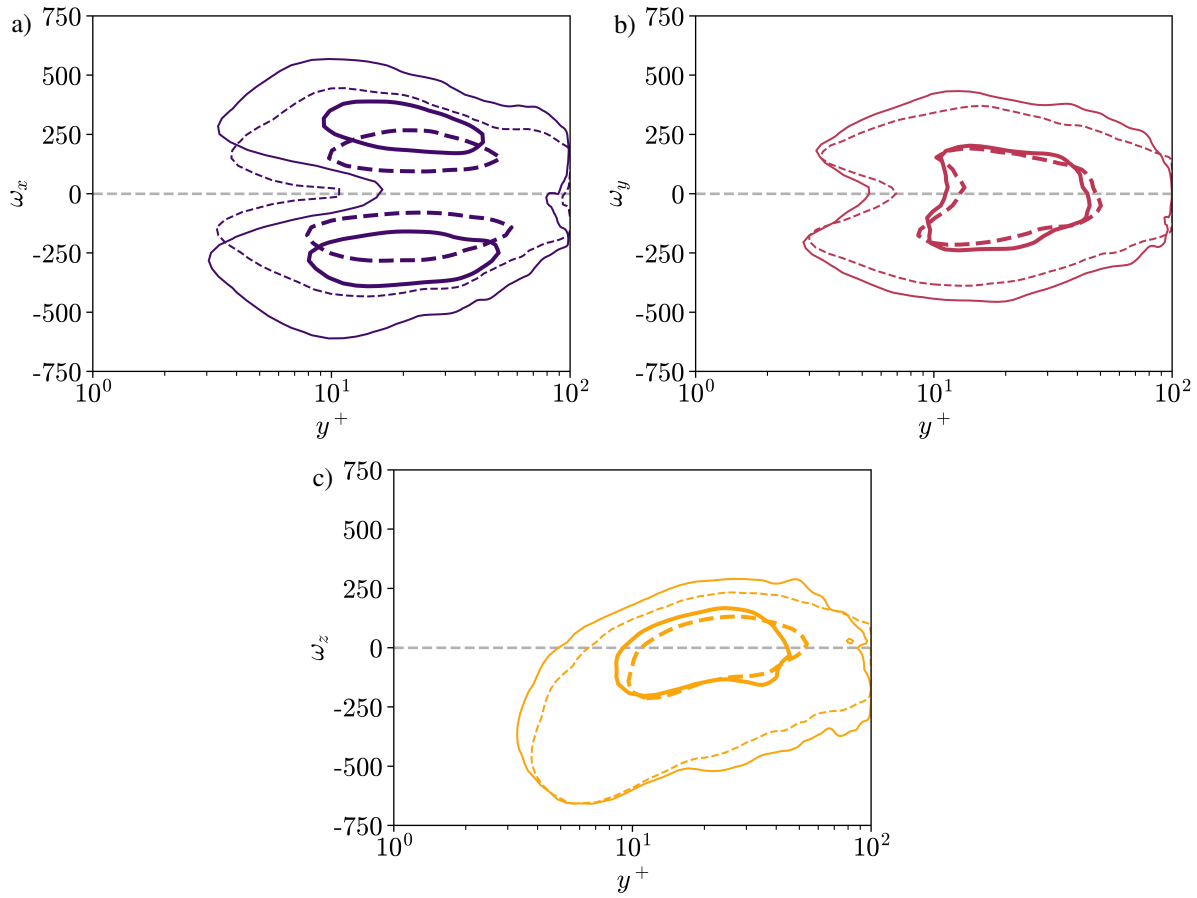


Figure 5.31: Joint $p.d.f(\omega_i^{C2_{\Omega'}}, y^+)$ (—) and joint $p.d.f(\omega_i^{\alpha_{perc,Q'}}, y^+)$ (- - -) for a) ω_x, b) ω_y, and c) ω_z. Contour levels contain 66% (thin line) and 33% (thick line) of the joint probability mass, respectively.

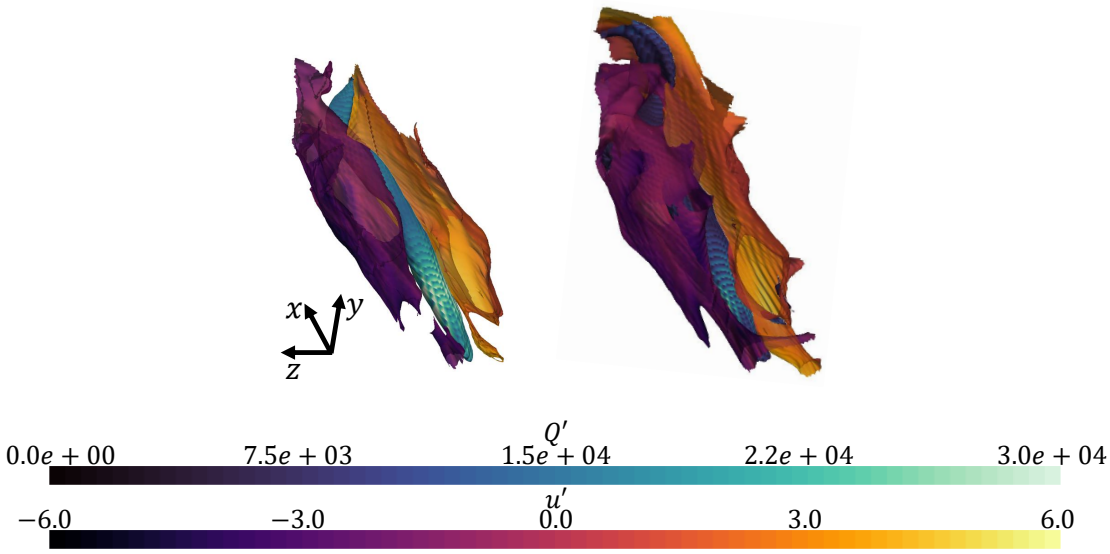


Figure 5.32: Instantaneous trimetric view of a near-wall ejection, sweep, and vortex collection as identified by the unsupervised method (left) and by percolation thresholds $\alpha^1_{perc,-u'v'}$ and $\alpha_{perc,Q'}$ (right).

minimum in figure 5.29(b).

5.4.2 Qualitative Comparison

The structures identified by thresholds $\alpha_{perc,Q'}$ and $\alpha^1_{perc,-u'v'}$ are compared against those found by the unsupervised method in section 5.2. Observing figure 5.32, the general shape of the structures remain unchanged. The intensity of Q' is lower in the vortex identified by percolation theory, and the vortex is slightly larger, suggesting a larger Q' decision boundary for the unsupervised method than $\alpha_{perc,Q'}$. These results are consistent with the statistical findings shown in figure 5.31. More cells further from the core of vortices are included in the statistics, which include larger populations of lesser vorticity that shift the vorticity distributions towards lower values. A similar result is found for the sweeps and ejections, whose structures are slightly larger and u' are slightly less intense, indicated by the darker colours in figure 5.32.

Percolation analysis was also applied to the Q'^s fields of section 5.3, following the process outlined earlier. A logarithmic threshold range of $\alpha = 10^{-4}, ..., 10^2$ with length 100 was used, with fields satisfying $Q'^s > \alpha$ being generated. V_{lar}/V_{tot} was computed for each threshold value, yielding the earlier shown figure 4.2 and $\alpha_{perc} \simeq 1.6$. The bounding surfaces of some arbitrary

thresholds, as well as the percolation threshold, are compared against the bounding surfaces found by the unsupervised method in figure 5.33. Perceptually, the unsupervised boundaries correspond to a threshold value just above α_{perc}, keeping in mind that the unsupervised "threshold" is more specifically a multivariate threshold (Q'^s & $-u'v'^s$).

5.4.3 Discussion

In comparing the unsupervised method with the percolation method, it is difficult to say which is better. Training times are comparable, where training time for the percolation method involves computing cluster volumes across an array of thresholds, and training for the unsupervised method involves the self-organization of weights within the SOM. Inference time for percolation theory is negligible, once the threshold is determined, inference only requires a binary operation, e.g., $Q(\boldsymbol{x}, t) > \alpha$; SOM inference, however, requires determining the BMU for each input, making inference more costly (see appendix B.1.1 for inference times).

A slight point of concern for the percolation method is how to interpret the plot. Whether one should select the location directly after a steep gradient, i.e., an "elbow" on the curve, as done by Del Álamo et al. [2006]; or the local minimum shortly after, as done by Motoori and Goto [2021], is not clear. A sharp gradient in the plot acts as a proxy for a sharp transition on the volumetric contours of the target flow variable, indicating the sudden increase of a population of similarly sized structures; a local minimum indicates a population of even more distributed evenly shaped structures, but not necessarily a rapid transition point. The user must make a choice of where on the graph the percolation threshold is likely to be, and if the choice is not obvious, more human bias is involved in the threshold selection process. The unsupervised learning method offers a "hands-free" alternative, where the choice is automatically made for the user.

Last, the percolation method obtains thresholds univariately. When structures from two dynamic features are compared, they overlap in space and time. There is no physical reason why multiple structures attributed to different features ought not to overlap, yet, discerning the dynamical interactions of the structures can be challenging when they do (e.g., see Robinson [1991, p. 317] and Motoori and Goto [2021, p. 6]). The unsupervised learning method obtains a multivariate threshold, and returns clusters that are mutually exclusive from one another. Observation of dynamical interactions can be easier to process in this case. Exclusive ownership of specific grid cells also enables the tracking of proportional statistics; e.g., for a transient process one could track the volumetric share of energy producing and dissipating structures, knowing that respective volumes belong to distinctly separate processes.

The percolation and unsupervised methods are both effective methods for objectively elucidating regions of interest. Percolation is simple and easy to apply, yet vulnerable to bias and limited

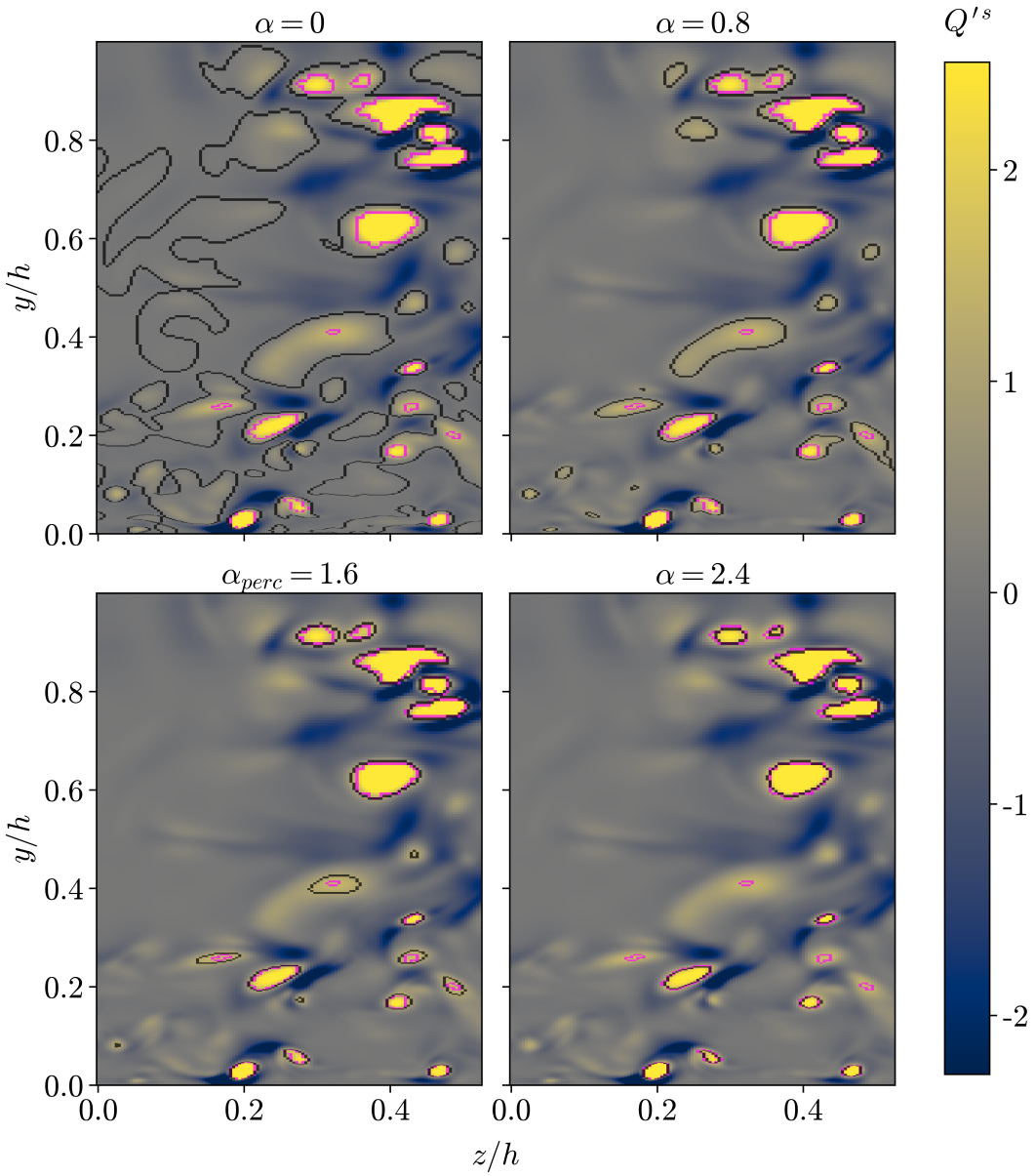


Figure 5.33: Instantaneous cross-stream snapshots of Q'^s overlaid with cluster boundaries using static thresholds (black) and the unsupervised learning method (pink).

to a univariate threshold; the unsupervised method is more complex, yet minimizes human bias and enables multivariate thresholding.

Chapter 6

Conclusions

The pursuit of this work was motivated by the desire to find an objective method to label regions of interest in turbulent flows to elucidate the beautiful structures that manifest in turbulent boundary layer flow. Leafing through pages of the boundary layer literature, one often comes across figures that contain three-dimensional visualizations, with an iso-surface threshold stated in the caption that leaves the reader wondering: Why was this particular value used? This work hopes to provide a solution to remedy this commonly encountered conundrum. An unsupervised learning method has been employed to identify CS at various scales in a turbulent open channel flow, offering a "hands-off" approach to label salient regions of flow. The batch SOM outlined in § 4 is used to obtain a low-dimensional representation of the input data, the low-dimensional map is further clustered using k-means, and additional steps are taken to ensure experimental reproducibility. Structures pertaining to the various wall-normal regions introduced in § 3 were investigated, those regions being the near-wall region ($y^+ \lessapprox 60$), wall-region ($y^+ \lessapprox 100$), and entire wall-normal domain. Section 5 shows that the resulting near-wall ejection, sweep, and vorticity clusters are consistent with boundary layer literature and comply with physical intuitions. Near-wall quasi-streamwise vortices flanked by ejections and sweeps are unveiled by the unsupervised clustering method (§ 5.1). Quadrant distributions of ejection and sweep clusters demonstrate stronger sweeps in the lower buffer layer ($y^+ \lessapprox 15$) and stronger ejections in the outer buffer and log layer ($15 \gtrapprox y^+ \lessapprox 100$) (§ 5.2). Sweep and ejection strength are found to be more balanced in the outer layer, where inhomogenieties introduced by the wall have a lesser effect. Vortices are found to be more streamwise-oriented in the near-wall; populations of transverse-oriented vortices are found in the log-layer; and vortices in the outer-layer are more structurally isotropic (§ 5.2). LSMs and VLSMs are observed in the outer layer, resembling outsized equivalents to the smaller scale ejection–vortex–sweep arrangements found in the wall-region (§ 5.2). Unsupervised clustering offers a promising alternative to manual filtering. Applicable to any numerical flow, unsupervised

clustering enables researchers across the realm of fluid dynamics to inspect salient regions of flow with increased objectivity.

Further directions of this work include to assess the generalizability of the unsupervised method and how its corresponding outputs change with Reynolds number or alternative flow configurations, such as stratified turbulence or a spatially developing flow. Another route could be to explore alternative unsupervised learning methods to reduce training and inference time; though, with the lack of ground truths the hurdle remains of how to assess model validity in an unsupervised setting. Last, the outer layer collections of ejecting and sweeping LSMs and VLSMs that were observed in § 5.3 may warrant further investigation to establish an understanding of their role in the self-sustaining process of wall-bounded turbulence.

References

R. J. Adrian. Hairpin vortex organization in wall turbulence. *Physics of Fluids*, 19(4):041301, 2007. doi: 10.1063/1.2717527. URL https://doi.org/10.1063/1.2717527.

R. K. Anand, B. J. Boersma, and A. Agrawal. Detection of turbulent/non-turbulent interface for an axisymmetric turbulent jet: Evaluation of known criteria and proposal of a new criterion. *Experiments in fluids*, 47(6):995–1007, 2009. ISSN 1432-1114.

A. Atoufi, K. A. Scott, and M. L. Waite. Wall turbulence response to surface cooling and formation of strongly stable stratified boundary layers. *Physics of fluids (1994)*, 31(8):85114–, 2019. ISSN 1070-6631.

A. Atoufi, K. A. Scott, and M. L. Waite. Characteristics of quasistationary near-wall turbulence subjected to strong stable stratification in open-channel flows. *Phys. Rev. Fluids*, 5:064603, Jun 2020. doi: 10.1103/PhysRevFluids.5.064603. URL https://link.aps.org/doi/10.1103/PhysRevFluids.5.064603.

A. M. Bagirov, R. M. Aliguliyev, and N. Sultanova. Finding compact and well-separated clusters: Clustering using silhouette coefficients. *Pattern Recognition*, 135:109144, 2023. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2022.109144. URL https://www.sciencedirect.com/science/article/pii/S0031320322006239.

J. W. Brooke and T. J. Hanratty. Origin of turbulence-producing eddies in a channel flow. *Physics of Fluids A: Fluid Dynamics*, 5(4):1011–1022, 1993. doi: 10.1063/1.858666. URL https://doi.org/10.1063/1.858666.

M. Bussov and J. Nättilä. Segmentation of turbulent computational fluid dynamics simulations with unsupervised ensemble learning. *CoRR*, abs/2109.01381, 2021. URL https://arxiv.org/abs/2109.01381.

R. Dawkins and J. Lenzová. *Flights of Fancy: Defying Gravity by Design and Evolution*. Head of Zeus, 2021. ISBN 9781838937874. URL https://books.google.ca/books?id=3CsOEAAAQBAJ.

J. C. Del Álamo, J. Jiménez, P. Zandonade, and R. D. Moser. Self-similar vortex clusters in the turbulent logarithmic region. *Journal of Fluid Mechanics*, 561:329–358, August 2006. doi: 10.1017/S0022112006000814.

S. Dong, A. Lozano-Durán, A. Sekimoto, and J. Jiménez. Coherent structures in statistically stationary homogeneous shear turbulence. *Journal of Fluid Mechanics*, 816:167–208, 2017. doi: 10.1017/jfm.2017.78.

P. A. Durbin and B. A. Pettersson Reif. *Statistical Theory and Modeling for Turbulent Flows*. Wiley, August 2010. doi: 10.1002/9780470972076. URL https://doi.org/10.1002/9780470972076.

K. Fukami, K. Fukagata, and K. Taira. Super-resolution reconstruction of turbulent flows with machine learning. *Journal of fluid mechanics*, 870:106–120, 2019. ISSN 0022-1120.

K. Fukami, K. Fukagata, and K. Taira. Machine learning based spatio-temporal super resolution reconstruction of turbulent flows. 2020.

M. Galbraith and M. Visbal. Implicit large eddy simulation of low reynolds number flow past the sd7003 airfoil. volume 225, 01 2008. ISBN 978-1-62410-128-1. doi: 10.2514/6.2008-225.

M. J. P. Hack and T. A. Zaki. Streak instabilities in boundary layers beneath free-stream turbulence. *Journal of fluid mechanics*, 741:280–315, 2014. ISSN 0022-1120.

P. He. A high order finite difference solver for massively parallel simulations of stably stratified turbulent channel flows. *Computers & Fluids*, 127:161–173, 2016. ISSN 0045-7930. doi: https://doi.org/10.1016/j.compfluid.2015.12.012. URL https://www.sciencedirect.com/science/article/pii/S0045793015004090.

M. R. Head and P. Bandyopadhyay. New aspects of turbulent boundary-layer structure. *Journal of fluid mechanics*, 107(1):297–338, 1981. ISSN 0022-1120.

T. B. Hedley and J. F. Keffer. Turbulent/non-turbulent decisions in an intermittent flow. *Journal of fluid mechanics*, 64(4):625–644, 1974. ISSN 0022-1120.

X. Hu, Z. Deng, J. Zhang, and W. Zhang. Aerodynamic behaviors in supersonic evacuated tube transportation with different train nose lengths. *International Journal of Heat*

and Mass Transfer, 183:122130, 2022. ISSN 0017-9310. doi: https://doi.org/10.1016/j.ijheatmasstransfer.2021.122130. URL https://www.sciencedirect.com/science/article/pii/S0017931021012369.

J. Hwang, J. Lee, H. J. Sung, and T. A. Zaki. Inner–outer interactions of large-scale structures in turbulent channel flow. *Journal of Fluid Mechanics*, 790:128–157, 2016. doi: 10.1017/jfm.2016.3.

E. Jagodinski, X. Zhu, and S. Verma. Uncovering dynamically critical regions in near-wall turbulence using 3d convolutional neural networks. 2020.

J. Jeong and F. Hussain. On the identification of a vortex. *Journal of fluid mechanics*, 285:69–94, 1995. ISSN 0022-1120.

J. Jeong, F. Hussain, W. Schoppa, and J. Kim. Coherent structures near the wall in a turbulent channel flow. *Journal of Fluid Mechanics*, 332:185–214, 1997. doi: 10.1017/S0022112096003965.

J. Jiménez. Coherent structures in wall-bounded turbulence. *Journal of Fluid Mechanics*, 842: P1, 2018. doi: 10.1017/jfm.2018.144.

A. M. Kalteh, P. Hjorth, and R. Berndtsson. Review of the self-organizing map (som) approach in water resources: Analysis, modelling and application. *Environmental modelling & software : with environment data news*, 23(7):835–845, 2008. ISSN 1364-8152.

H. Kim, J. Kim, S. Won, and C. Lee. Unsupervised deep learning for super-resolution reconstruction of turbulence. 2020.

J. Kim and C. Lee. Deep unsupervised learning of turbulence for inflow generation at various reynolds numbers. *Journal of computational physics*, 406:109216–, 2020. ISSN 0021-9991.

J. Kim and P. Moin. The structure of the vorticity field in turbulent channel flow. ii - study of ensemble-averaged fields. *Journal of fluid mechanics*, 162:339–363, 1986. ISSN 0022-1120.

K. Kim and P. Moin. The structure of the vorticity field in turbulent channel flow. i - analysis of instantaneous fields and statistical correlations. *Journal of fluid mechanics*, 155:441–464, 1985. ISSN 0022-1120.

S. J. Kline and P. W. Runstadler. Discussion: "some preliminary results of visual studies of the flow model of the wall layers of the turbulent boundary layer" (kline, s. j., and runstadler, p. w., 1959, asme j. appl. mech., 26, pp. 166–170). *Journal of applied mechanics*, 27(1):213–214, 1960. ISSN 0021-8936.

K. Koffka. *Principles of Gestalt psychology*. 1935.

T. Kohonen. The self-organizing map. *Neurocomputing (Amsterdam)*, 21(1):1–6, 1998. ISSN 0925-2312.

T. Kohonen. *Self-Organizing Maps*. Springer Berlin Heidelberg, 2001. doi: 10.1007/978-3-642-56927-2. URL https://doi.org/10.1007/978-3-642-56927-2.

T. Kohonen. Essentials of the self-organizing map. *Neural networks*, 37:52–65, 2013. ISSN 0893-6080.

P. Kundu. *Fluid Mechanics 2nd Edition*. Elsevier Exclusive, 2003. ISBN 9788178672434. URL https://books.google.ca/books?id=9MTCbwAACAAJ.

J. Lee, J. H. Lee, J. Choi, and H. J. Sung. Spatial organization of large- and very-large-scale motions in a turbulent channel flow. *Journal of fluid mechanics*, 749:818–840, 2014. ISSN 0022-1120.

K. E. Lilly. The representation of small-scale turbulence in numerical simulation experiments. 1966.

S. Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2): 129–137, 1982. ISSN 0018-9448.

A. Lozano-Durán, O. Flores, and J. Jiménez. The three-dimensional structure of momentum transfer in turbulent channels. *Journal of Fluid Mechanics*, 694:100–130, 2012. doi: 10.1017/jfm.2011.524.

Y. Motoori and S. Goto. Hierarchy of coherent structures and real-space energy transfer in turbulent channel flow. *Journal of Fluid Mechanics*, 911:A27, 2021. doi: 10.1017/jfm.2020.1025.

G. Narasimhan, C. Meneveau, and T. A. Zaki. Large eddy simulation of transitional channel flow using a machine learning classifier to distinguish laminar and turbulent regions. *Phys. Rev. Fluids*, 6:074608, Jul 2021. doi: 10.1103/PhysRevFluids.6.074608. URL https://link.aps.org/doi/10.1103/PhysRevFluids.6.074608.

T. Omar, A. Alzahrani, and M. Zohdy. Clustering approach for analyzing the student's efficiency and performance based on data. *J. data anal. inf. process.*, 08(03):171–182, 2020.

S. A. Orszag and G. S. Patterson. Numerical simulation of three-dimensional homogeneous isotropic turbulence. *Physical Review Letters*, 28(2):76–79, January 1972. doi: 10.1103/physrevlett.28.76. URL https://doi.org/10.1103/physrevlett.28.76.

U. Piomelli, W. H. Cabot, P. Moin, and S. Lee. Subgrid-scale backscatter in turbulent and transitional flows. *Physics of Fluids A: Fluid Dynamics*, 3(7):1766–1771, 1991. doi: 10.1063/1.857956. URL https://doi.org/10.1063/1.857956.

G. D. Portwood, S. M. de Bruyn Kops, J. R. Taylor, H. Salehipour, and C. P. Caulfield. Robust identification of dynamically distinct regions in stratified turbulence. *Journal of fluid mechanics*, 807, 2016. ISSN 0022-1120.

A. J. Richardson, C. Risien, and F. A. Shillington. Using self-organizing maps to identify patterns in satellite imagery. *Progress in oceanography*, 59(2-3):223–239, 2003. ISSN 0079-6611.

L. F. Richardson. *Weather Prediction by Numerical Process*. Cambridge Mathematical Library. Cambridge University Press, 1 edition, 1922. doi: 10.1017/CBO9780511618291.

S. K. Robinson. *The kinematics of turbulent boundary layer structure*. 1991.

Stephen K. Robinson. Coherent motions in the turbulent boundary layer. *Annual Review of Fluid Mechanics*, 23:601–639, January 1991. doi: 10.1146/annurev.fl.23.010191.003125.

P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987. ISSN 0377-0427.

E. H. Sloane. Reductionism. *Psychological Review*, 52(4):214–223, 1945. doi: 10.1037/h0059151.

P. R. Spalart. Direct simulation of a turbulent boundary layer up to r = 1410. *Journal of Fluid Mechanics*, 187:61–98, 1988. doi: 10.1017/S0022112088000345.

C. M. Ströfer, J. Wu, H. Xiao, and E. Paterson. Data-driven, physics-based feature extraction from fluid flow fields. 2018.

C. B. Sullivan and A. A. Kaszynski. Pyvista: 3d plotting and mesh analysis through a streamlined interface for the visualization toolkit (vtk). *Journal of Open Source Software*, 4(37):1450, 2019. doi: 10.21105/joss.01450. URL https://doi.org/10.21105/joss.01450.

R. L. Thorndike. Who belongs in the family? *Psychometrika*, 18(4):267–276, December 1953. doi: 10.1007/bf02289263. URL https://doi.org/10.1007/bf02289263.

T. Topór. An integrated workflow for micp-based rock typing: A case study of a tight-gas sandstone reservoir in the baltic basin (poland). *Nafta-Gaz*, 76:219–229, 04 2020. doi: 10.18668/NG.2020.04.01.

J. Vesanto and E. Alhoniemi. Clustering of the self-organizing map. *IEEE transactions on neural networks*, 11(3):586–600, 2000. ISSN 1045-9227.

F. Waleffe. On a self-sustaining process in shear flows. *Physics of Fluids*, 9(4):883–900, 1997. doi: 10.1063/1.869185. URL https://doi.org/10.1063/1.869185.

J. M. Wallace. Quadrant analysis in turbulence research: History and evolution. *Annual Review of Fluid Mechanics*, 48(1):131–158, 2016. doi: 10.1146/annurev-fluid-122414-034550. URL https://doi.org/10.1146/annurev-fluid-122414-034550.

J. M. Wallace, H. Eckelmann, and R. S. Brodkey. The wall region in turbulent shear flow. *Journal of fluid mechanics*, 54(1):39–48, 1972. ISSN 0022-1120.

W. W. Willmarth and S. S. Lu. Structure of the reynolds stress near the wall. *Journal of Fluid Mechanics*, 55(1):65–92, 1972. doi: 10.1017/S002211207200165X.

P. Wittek, S. C. Gao, I. S. Lim, and L. Zhao. somoclu: An efficient parallel library for self-organizing maps. *Journal of Statistical Software*, 78(9):1–21, 2017. doi: 10.18637/jss.v078.i09. URL https://www.jstatsoft.org/index.php/jss/article/view/v078i09.

A. A. Wray, P. Moin, and J. C. R. Hunt. Eddies, streams, and convergence zones in turbulent flows. 1988.

Z. Wu, J. Lee, C. Meneveau, and T. Zaki. Application of a self-organizing map to identify the turbulent-boundary-layer interface in a transitional flow. *Physical review fluids*, 4(2), 2019. ISSN 2469-990X.

J. C. Wyngaard. *Turbulence in the Atmosphere*. Cambridge University Press, 2010. doi: 10.1017/CBO9780511840524.

T. Yatsunenko, F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, A. P. Anokhin, A. C. Heath, B. Warner, J. Reeder, J. Kuczynski, J. G. Caporaso, C. A. Lozupone, C. Lauber, J. C. Clemente, D. Knights, R. Knight, and J. I. Gordon. Human gut microbiome viewed across age and geography. *Nature*, 486(7402):222–227, Jun 2012. ISSN 1476-4687. doi: 10.1038/nature11053. URL https://doi.org/10.1038/nature11053.

Z. Yuan, Z. Gu, Y. Wang, and X. Huang. Numerical investigation for the influence of the car underbody on aerodynamic force and flow structure evolution in crosswind. *Advances in Mechanical Engineering*, 10(10):1687814018797506, 2018. doi: 10.1177/1687814018797506. URL https://doi.org/10.1177/1687814018797506.

J. Zhou, R. J. Adrian, S. Balachandar, and T. M. Kendall. Mechanisms for generating coherent packets of hairpin vortices in channel flow. *Journal of Fluid Mechanics*, 387:353–396, 1999. doi: 10.1017/S002211209900467X.

APPENDICES

Appendix A

Nomenclature

Physical symbols

ρ	Density
ν	Kinematic viscosity
p	Pressure
h	Channel height
L_i	Channel length in direction i
x, y, z	Streamwise, wall-normal, and spanwise spatial directions
$\boldsymbol{u}$	Velocity
$\boldsymbol{u'}$	Fluctuating velocity
u, v, w	Streamwise, wall-normal, and spanwise velocity
u_i	Velocity component in direction i
$\overline{u_i}$ or U_i	Mean velocity in direction i
$\partial u_i/\partial x_j$ or $u_{i,j}$	Velocity gradient w.r.t direction j
Re	Reynolds Number
ω_i	Vorticity component i
ϵ_{ii}	Normal strain component ii
γ_{ij}	Shear strain component ij
P_{ij}	TKE production tensor

ε_{ij}	TKE dissipation tensor
T_{ij}	TKE transport tensor
D_{ij}	TKE diffusion tensor
Π_{ij}	TKE pressure work tensor
Q	Q-criterion
Ω_{ij}	Vorticity tensor
S_{ij}	Strain rate tensor
y^+	Vertical position in wall units
u^{C1_Ω}	Streamwise velocity in grid cells belonging to cluster $C1_\Omega$

Data sets

$\boldsymbol{D}$	Five-dimensional simulated flow dataset
$\boldsymbol{D}^s$	Standardized flow dataset
$\boldsymbol{X}$	Two-dimensional input dataset
$\boldsymbol{W}$	SOM weights map
$\boldsymbol{C}$	Clustered SOM map
$\boldsymbol{M}$	Four-dimensional mask dataset

Other symbols

α	Threshold value
$\boldsymbol{m}$	k-means model
$\boldsymbol{m}_{all}$	Batch of all k-means models
$\boldsymbol{n}_{all}$	Batch size of all k-means models
$\boldsymbol{m}_{best}$	Best k-means model within $\boldsymbol{m}_{all}$
k	Cluster count

k_{best}	Best cluster count
N_i	Node count i
Cx_{Ω}	Cluster from section 5.x characterized by Ω
DNS	Direct numerical simulation
LES	Large eddy simulation

Appendix B

Model Parameters

B.1 Framework parameters

B.1.1 SOM Parameters

A specification required for the SOM method is the size of the neural map. This map is to serve as a two-dimensional representation of the input data, and the shape of the map will limit the complexity of the further clustering step. Assessing the quality of the map is challenging due to the nature of unsupervised learning, i.e., one lacks a ground truth map to compare with. Finding an optimal map size thus incurs this challenge; yet, an attempt at moving towards "more-optimal" map dimensions is made here.

The criteria relevant to this optimization problem are minimizing training time while maintaining enough map resolution to make meaningful clusters. At the extrema of map resolution one can see that a $[N, M] = [1, 2]$ map limits analysis to only two clusters, whereas a highly

N	M	Training time (hrs)	Inference time (hrs)
5	5	7.3	0.94
10	10	14.0	1.16
20	20	29.2	1.89

Table B.1: Relationship between SOM map size and time to train. Inference times were each recorded for inference applied to a domain of shape $[N_t, N_x, N_y, N_z] = [1, 768, 384, 768]$.

resolved map, e.g., $[N, M] = [100, 100]$, is considerably large if one only expects a handful of clusters (as in the case of this study).

Table B.1 shows that increasing map resolution increases training time, with a $4x$ increase in map resolution corresponding to a $2x$ increase in train time. Inference time also increases, as there exist more map nodes in which to search for the BMU of each input. The trained weight maps using the three SOM resolutions are shown in B.1.1. Figure B.2 shows that there exists slight variation in how Q' is distributed within the three different map sizes. Referring to figures B.1.1 and B.2, as map resolution increases, so does refinement in the clusters, i.e., non-salient population increases and the other clusters become more concentrated. This is displayed perceptually in figure B.4, where cluster boundaries constrict with higher SOM resolution. The $[10, 10]$ map is chosen for this study as it provides rapid training and inference, while still providing refined clusters.

B.1.2 Ensuring reproducibility of k-means

The k-means algorithm is stochastic, sensitive to initialized centroid locations. A way to address this issue is to instantiate a batch of models, and select the best model of that batch ($\boldsymbol{m}_{best}$) based on equation 4.7. Figure B.5 shows that with a batch of $n_{all} = 1000$, all models return identical maps (in terms of cluster positions, specific colours are unimportant as they have yet to be labeled), signalling that this batch size ensures reproducibility of the k-means portion of the unsupervised framework.

B.1.3 Determining cluster count

The two data-driven criteria used to determine cluster count are the silhouette score and the intra-class variance. Silhouette coefficients are obtained for each cluster count, and the best cluster count is taken to be the count with the maximal silhouette coefficient. This is visualized in figures B.6, B.7, and B.8, where the maximum average silhouette coefficient (figure B.6(b)) for the SOM map obtained in section 5.1 was obtained with a cluster count of $k = 3$. Meanwhile, the intra-class variance for each cluster count is also obtained, as shown in B.9, in which the elbow method returns a cluster count of $k = 7$. The rounded average between the two cluster counts, $k = 3.5 \rightarrow 4$, is used for k_{best}.

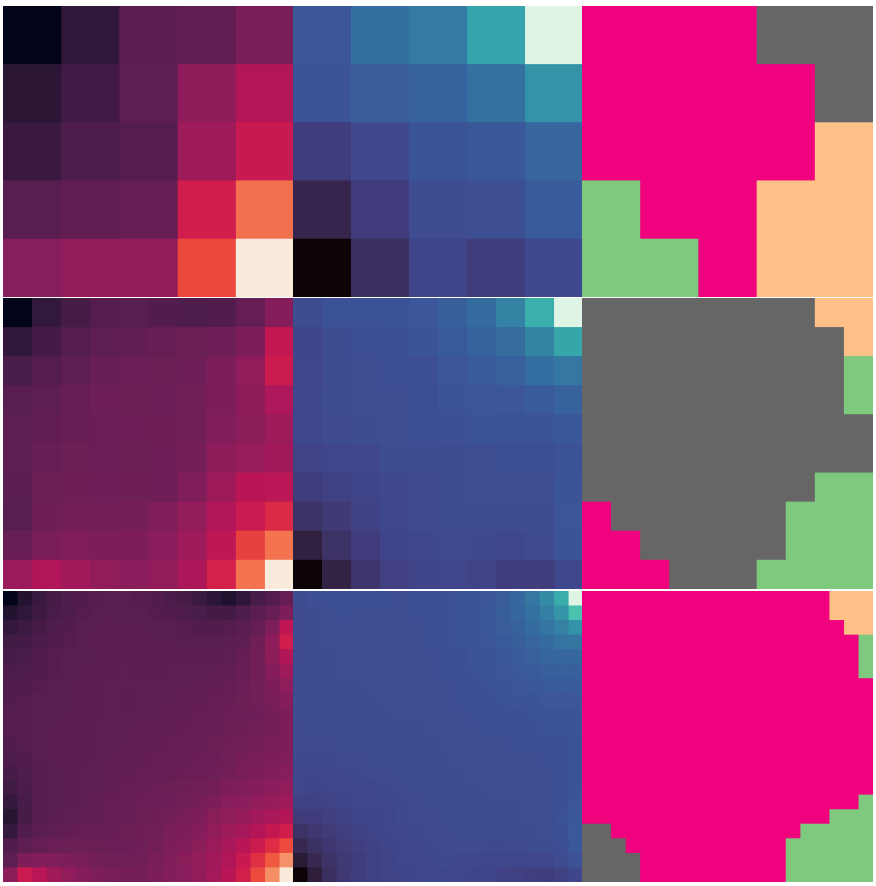


Figure B.1:
, $[10, 10]$, and $[20, 20]$ SOM map.]$\boldsymbol{W}_{-u'v'}$ (left), $\boldsymbol{W}_{Q'}$ (center), and $\boldsymbol{C2}$ (right) for a $[5, 5]$ (top), $[10, 10]$ (middle), and $[20, 20]$ (bottom) SOM map.

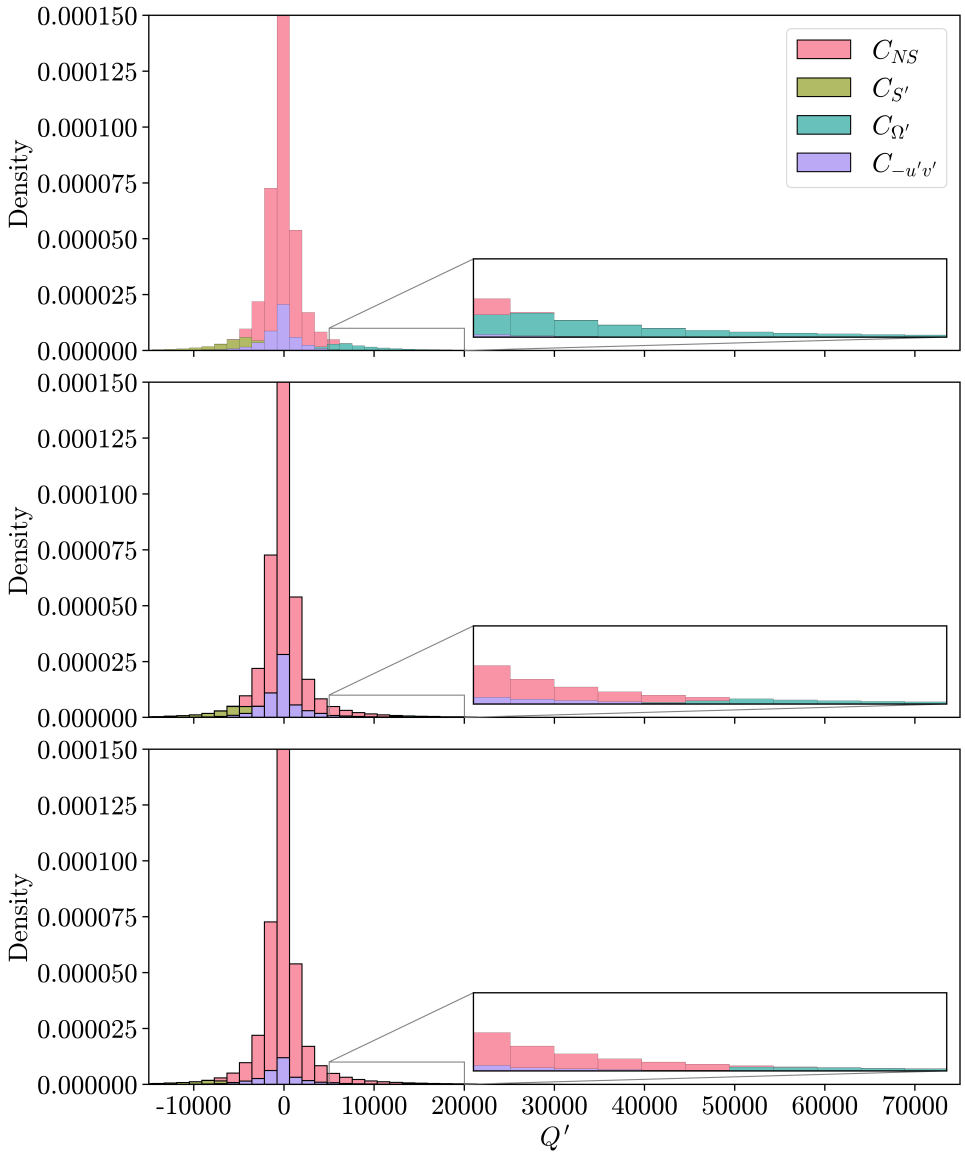


Figure B.2:
, [10, 10], and [20, 20] SOM map.]Stacked histogram of Q' within each cluster for a [5, 5] (top), [10, 10] (middle), and [20, 20] (bottom) SOM map.

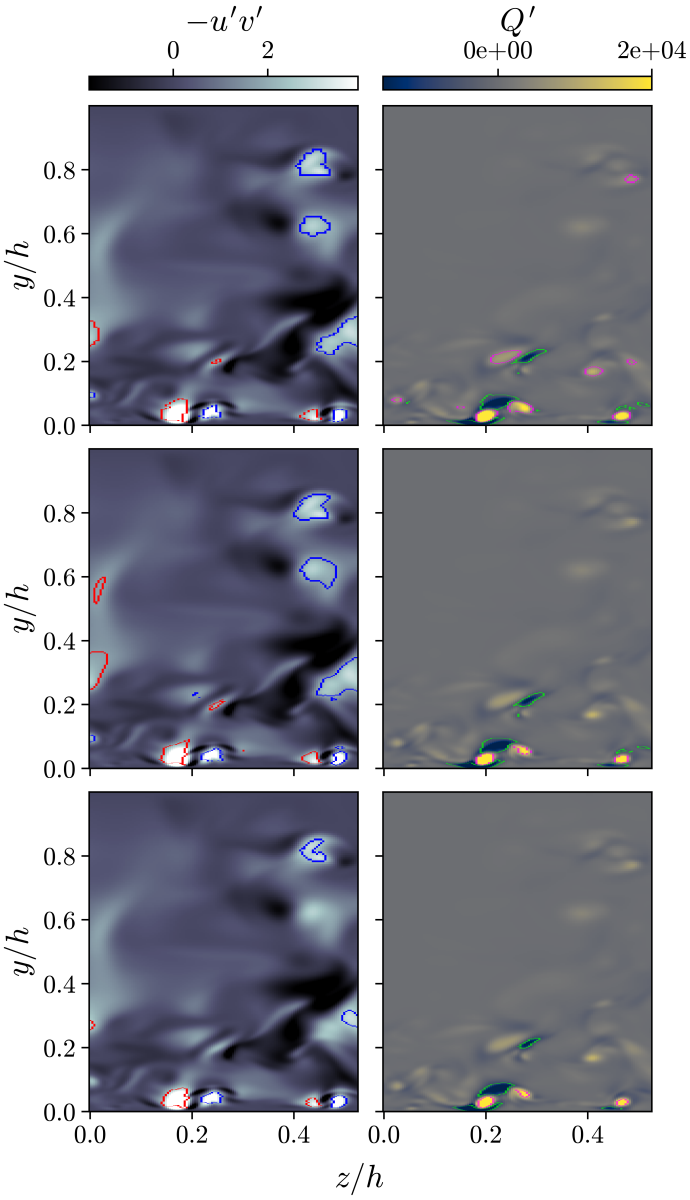


Figure B.3:
, [10, 10], and [20, 20] SOM map.]Instantaneous cross-stream snapshots of $-u'v'$ (left) and Q' (right) overlaid with cluster boundaries using a [5, 5] (top), [10, 10] (middle), and [20, 20] (bottom) SOM map. Red ($C2_{Q4}$), blue ($C2_{Q2}$), pink ($C2_{\Omega'}$) and green ($C2_{S'}$).

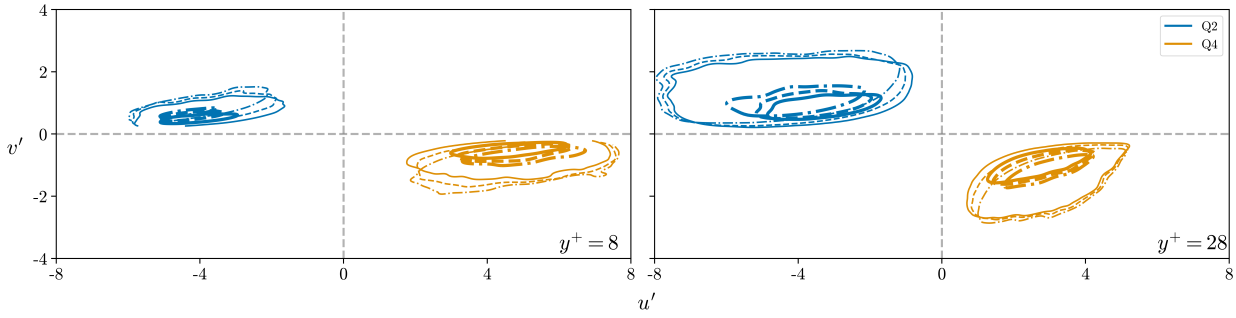


Figure B.4:
, [10, 10], and [20, 20].]Joint $p.d.f(u', v')$ of clusters $C2_{Q2}$ and $C2_{Q4}$ in the (left) lower buffer layer and (right) upper buffer layer for SOM maps of size [5, 5] (- - -); [10, 10] (——); and [20, 20] (- · -). Contour levels contain 66% (thin line) and 33% (thick line) of the joint probability mass, respectively.

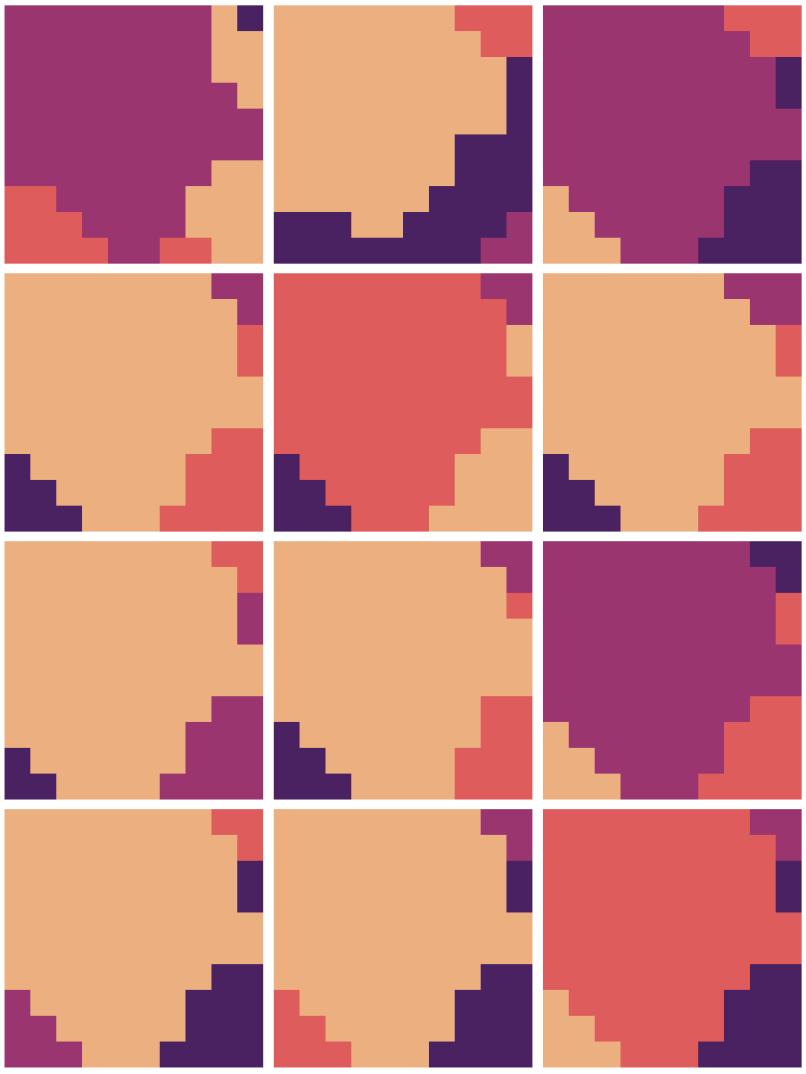


Figure B.5: Clustered maps $\boldsymbol{m}_{best}$ with batch size n_{all} = 1 (top), 10 (top-middle), 100 (bottom-middle), 1000 (bottom). Columns represent three arbitrary $\boldsymbol{m}_{best}$ using these batch sizes. More similar maps within a row indicate a more deterministic k-means model. Specific colours are unimportant as the maps are yet labeled.

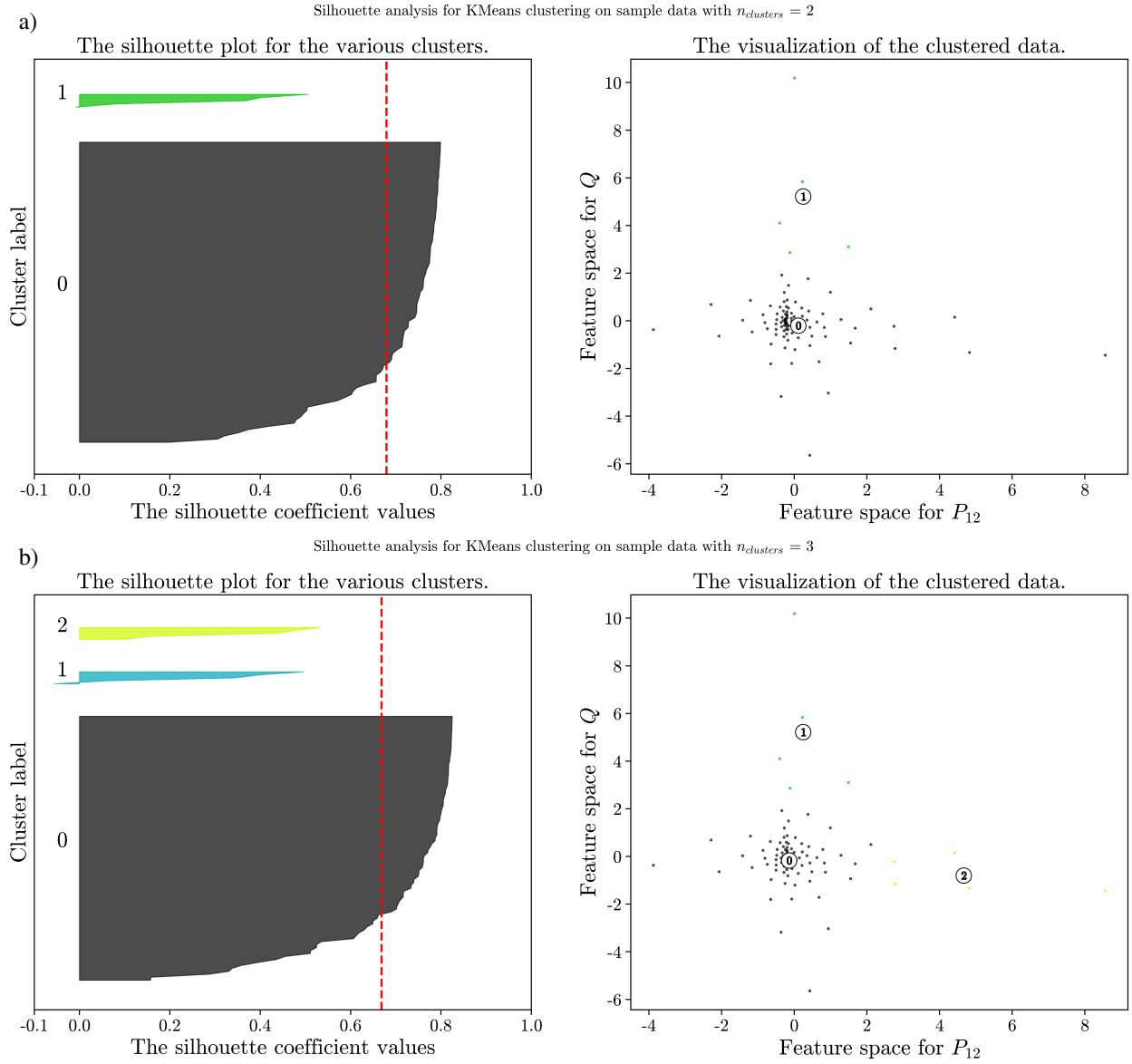


Figure B.6: Silhouette analysis for clusters a) $k = 2$ and b) $k = 3$. Red dashed line indicates the average silhouette coefficient. The left plot indicates the silhouette scores for each SOM weight within their respective cluster. The right plot shows how clusters are distributed in the SOM weight feature space.

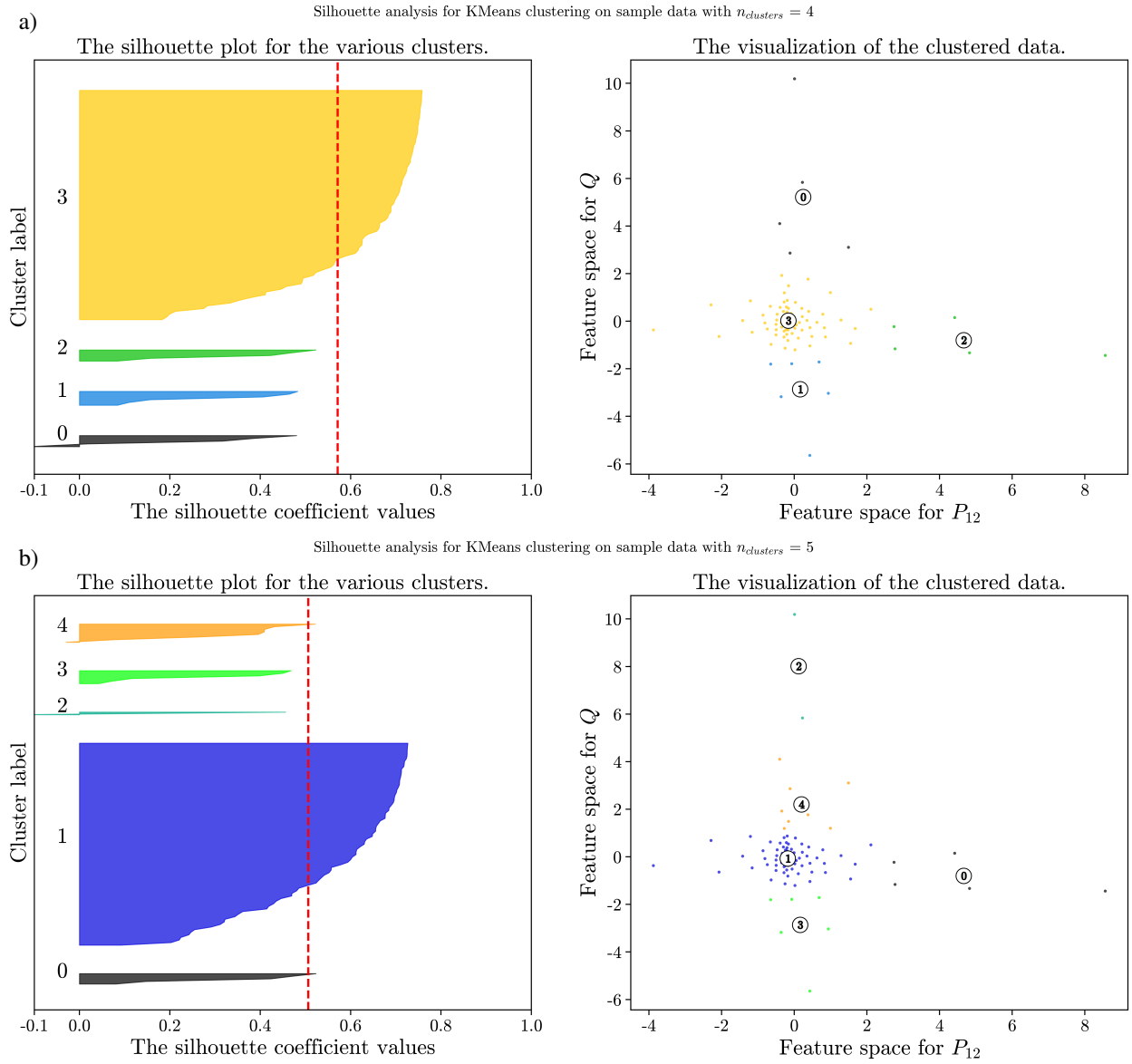


Figure B.7: Silhouette analysis for clusters a) $k = 4$ and b) $k = 5$. Red dashed line indicates the average silhouette coefficient.

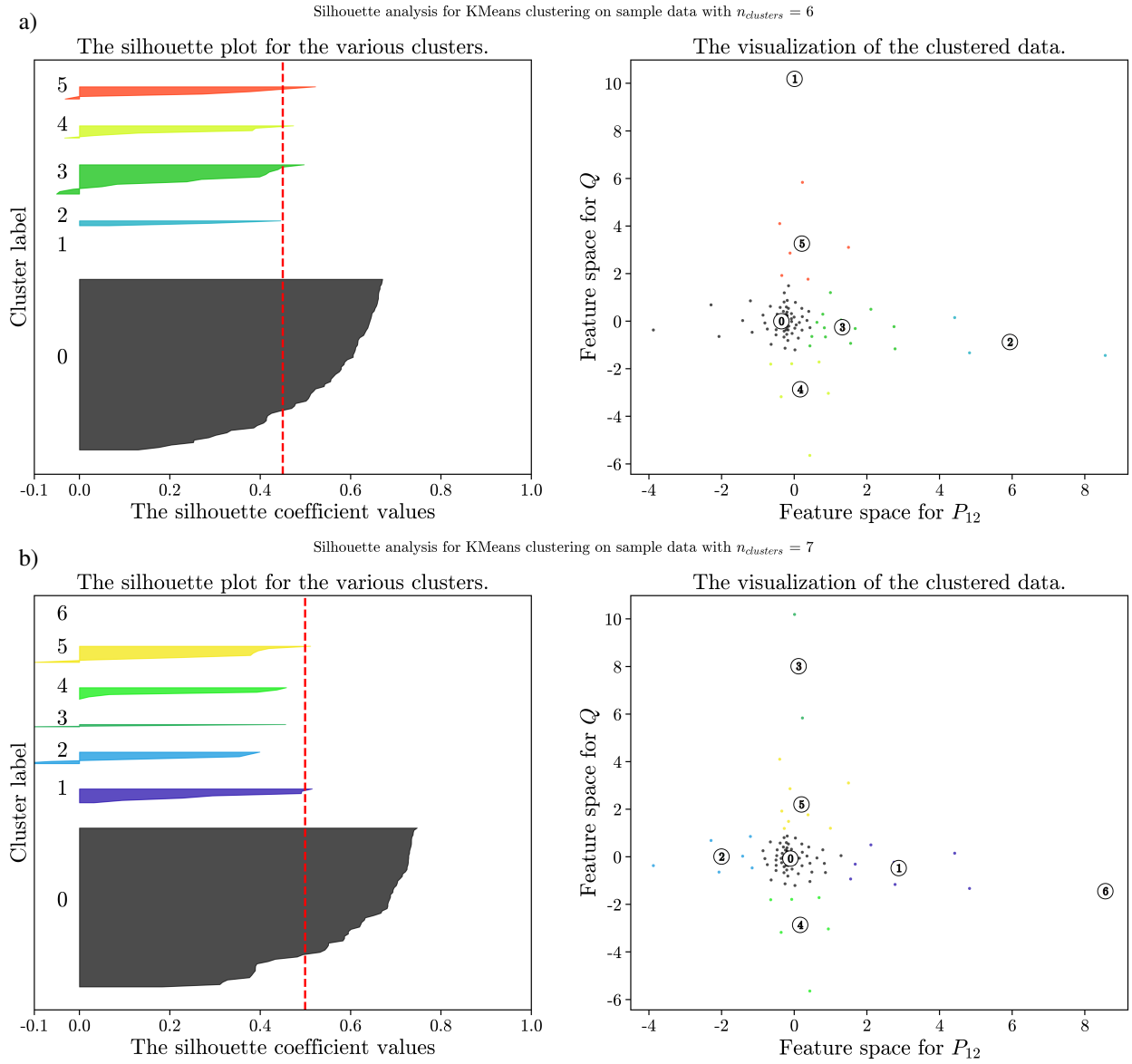


Figure B.8: Silhouette analysis for clusters a) $k = 6$ and b) $k = 7$. Red dashed line indicates the average silhouette coefficient.

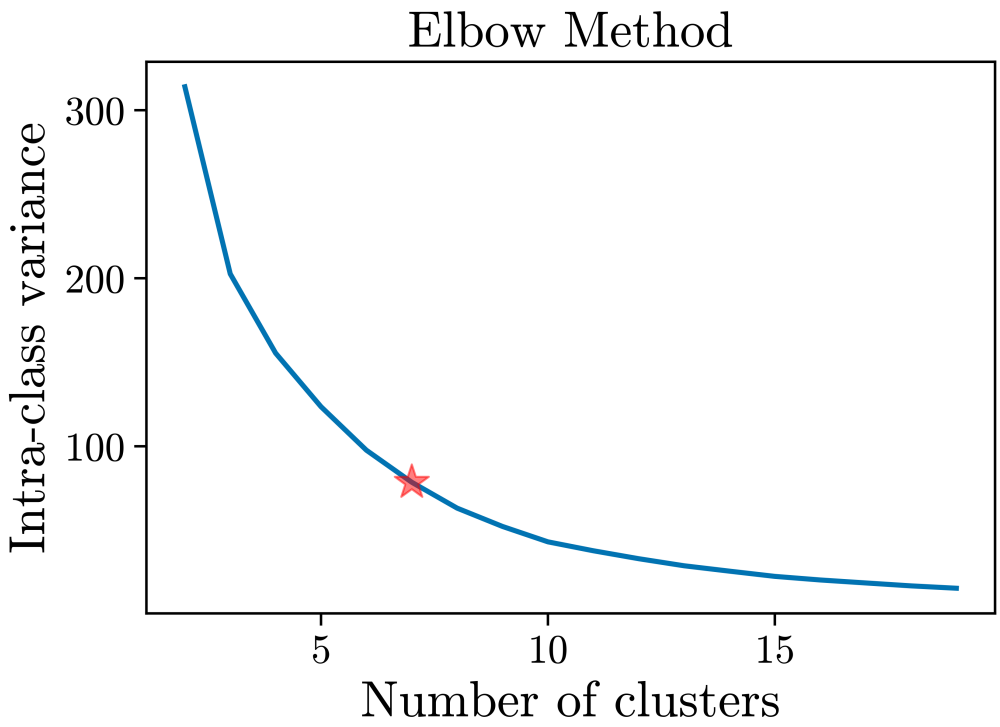


Figure B.9: Plot of intra-class variance against cluster count. The point on the plot with maximum curvature (the elbow) is indicated by ⋆.

Appendix C

Maths

C.1 Q-criterion

Vortices have long been known to play an important role in turbulent fluid dynamics. Consequently, a working definition of vortices was required. Prior to the introduction of more recent sophisticated vorticity metrics, vorticity magnitude, $|\omega|$, was used in attempt to identify vortices. Though, this method has no way of distinguishing between rotation from shear and rotation from swirling. Pressure minima was also used as a condition. However, while the core of a vortex necessarily has a pressure minimum, not every local pressure minimum represents a vortex core. Regions of low pressure indicate regions of accelerating fluid, not all of which occur in vortex cores.

The *Q-criterion*, Q, addresses these pitfalls. Q represents the local balance between the vorticity rate and strain rate of a fluid element, with large positive values of Q representing regions of strong rotation and negligible deformation. Q, the second invariant of the velocity gradient tensor, $\boldsymbol{\nabla u}$, can be written as

$$Q = \frac{1}{2}\left(||\boldsymbol{\Omega}||^2 - ||\boldsymbol{S}||^2\right) \tag{C.1}$$

where $||\boldsymbol{\Omega}|| = \left[\mathrm{tr}\left(\boldsymbol{\Omega}\boldsymbol{\Omega}^T\right)\right]^{1/2}$ and $||\boldsymbol{S}|| = \left[\mathrm{tr}\left(\boldsymbol{S}\boldsymbol{S}^T\right)\right]^{1/2}$. The antisymmetric component of $\boldsymbol{\nabla u}$, $\Omega_{ij} = \frac{1}{2}\left(u_{i,j} - u_{j,i}\right)$, represents local rotation, and the symmetric component of $\boldsymbol{\nabla u}$, $S_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right)$, represents local strain deformation.

Outlined below are the mathematical steps to represent C.1 in component form. For brevity,

derivative components will be written with just a subscript, e.g, instead of $u_{,y}$, u_y will be used. First, $\text{tr}\left(\boldsymbol{\Omega}\boldsymbol{\Omega}^T\right)$ can be rewritten as follows

$$
\begin{aligned}
||\boldsymbol{\Omega}|| &= \left[\text{tr}\left(\boldsymbol{\Omega}\boldsymbol{\Omega}^T\right)\right]^{1/2} \\
&= \left[\text{tr}\left(\frac{1}{2}\begin{bmatrix} 0 & u_y - v_x & u_z - w_x \\ v_x - u_y & 0 & v_z - w_y \\ w_x - u_z & w_y - v_z & 0 \end{bmatrix}\frac{1}{2}\begin{bmatrix} 0 & v_x - u_y & w_x - u_z \\ u_y - v_x & 0 & w_y - v_z \\ u_z - w_x & v_z - w_y & 0 \end{bmatrix}\right)\right]^{1/2} \\
&= \left[\frac{1}{4}\text{tr}\left(\begin{bmatrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{bmatrix}\begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}\right)\right]^{1/2} \\
&= \left[\frac{1}{4}\text{tr}\left(\begin{bmatrix} \omega_x^2 + \omega_y^2 & -\omega_y\omega_z & -\omega_x\omega_z \\ -\omega_y\omega_z & \omega_x^2 + \omega_z^2 & -\omega_x\omega_y \\ -\omega_x\omega_z & -\omega_x\omega_y & \omega_y^2 + \omega_z^2 \end{bmatrix}\right)\right]^{1/2} \\
&= \left[\frac{1}{4}\left(2\omega_x^2 + 2\omega_y^2 + 2\omega_z^2\right)\right]^{1/2} \\
&= \left[\frac{1}{2}\left(\omega_x^2 + \omega_y^2 + \omega_z^2\right)\right]^{1/2}.
\end{aligned}
$$

Repeating with $\text{tr}\left(\boldsymbol{S}\boldsymbol{S}^T\right)$,

$$
\begin{aligned}
||\boldsymbol{S}|| &= \left[\text{tr}\left(\boldsymbol{S}\boldsymbol{S}^T\right)\right]^{1/2} \\
&= \left[\text{tr}\left(\frac{1}{2}\begin{bmatrix} 2u_x & u_y + v_x & u_z + w_x \\ v_x + u_y & 2v_y & v_z + w_y \\ w_x + u_z & w_y + v_z & 2w_z \end{bmatrix}\frac{1}{2}\begin{bmatrix} 2u_x & v_x + u_y & w_x + u_z \\ u_y + v_x & 2v_y & w_y + v_z \\ u_z + w_x & v_z + w_y & 2w_z \end{bmatrix}\right)\right]^{1/2} \\
&= \left[\frac{1}{4}\text{tr}\left(\begin{bmatrix} \epsilon_{xx} & \gamma_{yx} & \gamma_{xz} \\ \gamma_{yx} & \epsilon_{yy} & \gamma_{yz} \\ \gamma_{xz} & \gamma_{yz} & \epsilon_{zz} \end{bmatrix}\begin{bmatrix} \epsilon_{xx} & \gamma_{yx} & \gamma_{xz} \\ \gamma_{yx} & \epsilon_{yy} & \gamma_{yz} \\ \gamma_{xz} & \gamma_{yz} & \epsilon_{zz} \end{bmatrix}\right)\right]^{1/2} \\
&= \left[\frac{1}{4}\text{tr}\left(\begin{bmatrix} \epsilon_{xx}^2 + \gamma_{xy}^2 + \gamma_{xz}^2 & \cdots & \cdots \\ \cdots & \gamma_{xy}^2 + \epsilon_{yy}^2 + \gamma_{yz}^2 & \cdots \\ \cdots & \cdots & \gamma_{xz}^2 + \gamma_{yz}^2 + \epsilon_{zz}^2 \end{bmatrix}\right)\right]^{1/2}
\end{aligned}
$$

$$= \left[\frac{1}{4} \left(\left(\epsilon_{xx}^2 + \gamma_{xy}^2 + \gamma_{xz}^2 \right) + \left(\gamma_{xy}^2 + \epsilon_{yy}^2 + \gamma_{yz}^2 \right) + \left(\gamma_{xz}^2 + \gamma_{yz}^2 + \epsilon_{zz}^2 \right) \right) \right]^{1/2}$$
$$= \left[\frac{1}{2} \left(\gamma_{xy}^2 + \gamma_{xz}^2 + \gamma_{yz}^2 + \frac{1}{2} \left(\epsilon_{xx}^2 + \epsilon_{yy}^2 + \epsilon_{zz}^2 \right) \right) \right]^{1/2}.$$

Thus, the component form of Q can be written as

$$\begin{aligned} Q =& \frac{1}{2} \left(||\boldsymbol{\Omega}||^2 - ||\boldsymbol{S}||^2 \right) \\ =& \frac{1}{2} \left(\left(\left[\frac{1}{2} \left(\omega_x^2 + \omega_y^2 + \omega_z^2 \right) \right]^{1/2} \right)^2 - \left(\left[\frac{1}{2} \left(\gamma_{xy}^2 + \gamma_{xz}^2 + \gamma_{yz}^2 + \frac{1}{2} \left(\epsilon_{xx}^2 + \epsilon_{yy}^2 + \epsilon_{zz}^2 \right) \right) \right]^{1/2} \right)^2 \right) \\ =& \frac{1}{4} \left(\underbrace{\left(\omega_x^2 + \omega_y^2 + \omega_z^2 \right)}_{\text{rotation}} - \underbrace{\left[\gamma_{xy}^2 + \gamma_{xz}^2 + \gamma_{yz}^2 + \frac{1}{2} \left(\epsilon_{xx}^2 + \epsilon_{yy}^2 + \epsilon_{zz}^2 \right) \right]}_{\text{deformation}} \right). \end{aligned}$$