POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Identifying regions of importance in turbulent flows: an explainable deep learning approach

Tesi di Laurea Magistrale in
Aeronautical Engineering - Ingegneria Aeronautica

Giovanni Orazi, 11013201

Advisor:
Prof. Alessandro Chiarini

Co-advisors:
Prof. Alexander Stroh

Academic year:
2025-2026

Abstract: This work investigates the identification of dynamically important regions in turbulent channel flows using explainable deep learning, with a particular focus on the interaction between momentum transport and passive scalar dynamics. A three-dimensional U-Net convolutional neural network is trained on direct numerical simulation data to predict the short-time evolution of turbulent flow fields. Explainability is introduced through the Gradient-SHAP framework, which assigns an importance value to each grid point based on its contribution to the model prediction error. These importance fields are used to define SHAP-based structures via a percolation procedure, enabling an objective, data-driven characterization of flow regions that most influence temporal evolution. The framework is first validated on a canonical smooth channel flow by reproducing previously reported results and demonstrating low prediction errors and consistent structural statistics. The methodology is then extended to a turbulent channel flow with temperature treated as a passive scalar. Multiple definitions of SHAP-based structures are examined to assess how different loss formulations influence the identification of dynamically relevant regions. The resulting structures are compared with classical coherent structures defined by Reynolds-stress and scalar-flux criteria through joint probability density analyses. The results show that SHAP-based structures partially overlap with traditional coherent structures while highlighting regions less emphasized by purely kinematic criteria. In the passive scalar case, the analysis highlights the asymmetry in causal influence between velocity and temperature fields and demonstrates that temperature-focused SHAP formulations are necessary to isolate thermally relevant structures. Overall, the study establishes an explainable, data-driven framework for identifying influential turbulent structures and provides practical guidelines for its application to coupled momentum–scalar transport problems.

Key-words: Turbulence, Machine learning, Shapley values, Explainability, Coherent structures

1. Introduction

For hundreds of years, countless scientists and engineers have attempted to tackle the seemingly insurmountable complexity of disorderly flows, characterized by their inherent multiscale and nonlinear nature.

Many prolific and renowned scientists, such as Euler, Navier, Stokes, Boussinesq, Poisson, and Cauchy, took an interest in fluid dynamics and established its theoretical foundation.

While the majority of the progress in turbulence theory in its early years was achieved from purely empirical observations, a significant theoretical breakthrough was reached by the Russian physicist Andrej Nikolaevič Kolmogorov, building upon the statistical work of Taylor [38, 39] and von Kármán [23] and the idea of energy cascade first introduced by Richardson [32]. Through his investigation, Kolmogorov reached many theoretical results regarding the characteristic scales of turbulent flows, the form of the structure functions, and the energy transfer rate in homogeneous isotropic turbulence. While the results showed good agreement with many experiments, subsequent research identified limitations in Kolmogorov's original phenomenology, particularly regarding internal intermittency and small-scale fluctuations [10, 21]. Nonetheless, Kolmogorov's legacy and contributions remain the bedrock on top of which turbulent research flourished and developed.

With the exponential increase of computational power in the 1980s, a new powerful tool became available to turbulence researchers, while until then, progress was made only through experimental data and theoretical considerations, the numerical integration of the Navier-Stokes equations opened the path to performing simulations of flows of interest. Kim et al. [18] simulated the simplest complete example of a wall-bounded flow, a turbulent channel. They performed a Direct Numerical Simulation (DNS), where all the spatial and temporal scales of the flow are resolved. Through a computational approach, a vast amount of high-quality data becomes available, allowing for a complete characterization of the simulated flow across all of its scales. The transition from the statistical era of Kolmogorov to the computational era of DNS shifted the focus from universal scaling laws to the spatial organization of the flow. Despite the wealth of data provided by numerical integration, the challenge remains in the objective extraction of information from the resulting chaotic fields. This necessity gave rise to the coherent-structure paradigm, which seeks to simplify the high-dimensional complexity of the Navier-Stokes equations into a reduced-order description of organized fluid motions.

The coherent-structure paradigm did not arise abruptly with the advent of DNS, but rather evolved from earlier experimental observations that revealed unexpectedly organized motions within turbulent flows. Although Kolmogorov's statistical framework [22] successfully characterized inertial-range universality, detailed measurements in wall-bounded turbulence during the 1960s began to expose persistent spatial patterns incompatible with a purely random picture. Landmark flow-visualization experiments demonstrated the presence of elongated low-speed streaks and quasi-streamwise vortices in the near-wall region [19, 20]. These findings suggested that turbulence contains dynamically significant structures responsible for momentum transfer and energy production.

Subsequent quantitative analyses strengthened this interpretation. The identification of bursting events and the introduction of quadrant analysis linked organized motions directly to Reynolds-stress production [17, 25]. With the advent of DNS, these structures were no longer inferred indirectly but observed in complete spatio-temporal detail. DNS confirmed that near-wall turbulence is sustained by a regeneration cycle involving streamwise vortices and streak instabilities, providing a concrete dynamical picture underlying the statistical observables.

This structural viewpoint was further reinforced by the emergence of a dynamical-systems interpretation of turbulence. Instead of viewing coherent structures merely as kinematic patterns, researchers began identifying exact invariant solutions of the Navier–Stokes equations that resemble experimentally observed motions. Finite-amplitude steady and traveling-wave solutions were computed in canonical shear flows [30], and later interpreted as building blocks of turbulence organized around unstable invariant sets in phase space [42]. The self-sustaining process (SSP) framework provided a mechanistic explanation for the cyclic interaction between vortices and streaks, linking coherent structures directly to nonlinear dynamics.

In parallel, objective methods for identifying coherent motions were developed. Eulerian vortex-detection criteria such as the Q-criterion and λ_2 method [14] allowed systematic extraction of vortical structures from velocity-gradient fields. Lagrangian coherent structure (LCS) theory later generalized the concept by identifying material transport barriers in unsteady flows through finite-time dynamical systems analysis. Complementary to these approaches, modal decompositions such as Proper Orthogonal Decomposition (POD), Dynamic Mode Decomposition (DMD), and resolvent analysis enabled reduced-order representations of dominant coherent motions and their amplification mechanisms [26].

Recent work has begun to augment the classical coherent-structure paradigm with explainable, data-driven methodologies aimed at objectively quantifying the dynamical relevance of different flow regions. In particular, the studies of Cremades et al. [6, 7] introduce a highly promising framework for investigating wall-bounded turbulence through explainable artificial intelligence (XAI). Their approach integrates U-Net deep convolutional neural networks (CNNs), capable of capturing multi-scale spatial correlations in turbulent fields [12], with the Shapley Additive Explanations (SHAP) algorithm [27], thereby enabling a quantitative attribution of predictive relevance to specific flow features.

In [6], SHAP values were assigned to pre-identified coherent structures, allowing classical events to be ranked according to their contribution to short-time flow predictions. This represents a major methodological advance: for the first time, the dynamical importance of coherent structures is assessed through a causal, model-based metric rather than inferred from kinematic indicators such as Reynolds stress alone. The results demonstrate

that the structures traditionally considered dominant are not always those most influential for the temporal evolution of the flow, challenging long-standing assumptions in wall-bounded turbulence research. The subsequent work [7] further strengthens the framework by removing the dependence on predefined structures. By computing importance directly at the grid-point level using the Gradient-SHAP algorithm, the authors construct spatial importance maps from which SHAP-based structures are systematically extracted. Comparison with classical coherent structures, such as Q events, streaks, and vortices, reveals only partial overlap, indicating that established kinematic criteria capture only a subset of the regions most relevant to flow evolution. This extension demonstrates the robustness, scalability, and discovery potential of the methodology, positioning it as a powerful tool for uncovering dynamically meaningful structures beyond traditional definitions.
Together, these studies establish XAI not merely as a predictive tool, but as a physically interpretable framework capable of reshaping how coherent structures are defined, ranked, and understood in wall-bounded turbulence. The primary objective of this research is to extend the XAI-based approach to turbulent channel flows from the 2025 work [7], with a specific focus on incorporating temperature as a passive scalar [16, 44]. By leveraging the Gradient-SHAP methodology, the study establishes a model-base importance index that characterizes flow regions based on their functional contribution to the evolution of the thermal field, rather than relying on traditional kinematic markers. This approach would allow for the objective isolation of high-importance thermal structures, independent of traditional definitions of coherent structures, thereby examining the possibility of subtle associations between velocity fluctuations and temperature that could be overlooked by classical approaches. Furthermore, such an investigation would enable a direct comparison between the spatial characteristics of momentum-driven and thermal-driven structures, providing deeper physical insight into the mechanisms of passive scalar transport in wall-bounded turbulence. Ultimately, the precise identification of these SHAP structures for the thermal field could facilitate the development of more efficient and targeted strategies for the control of heat dissipation and transport. Due to the novelty of the framework employed, this study seeks to determine the conditions required for extracting informative SHAP structures; since, to the author's knowledge, this is the first application of the framework to a channel flow with temperature, the aim is to provide specific guidelines for the use of the framework to this new case study.
The remainder of this work is organized as follows. In Section 2, we detail the numerical framework and the tools employed in this study, with particular emphasis on the DNS setup, the numerical solver, the architecture of the 3D U-Net model, the training process, and finally the Gradient-SHAP algorithm and the identification of the structures. Section 3 presents a comprehensive discussion of the results, providing a comparative analysis between classical structures and the objectively identified SHAP thermal structures. Finally, Section 4 summarizes the principal findings, and suggests potential avenues for future research into data-driven fluid mechanics.

2. Methodology

In the present section, the methods and tools employed in this work are described. Two simulations were carried out and analyzed in this work: the first one consists of a channel flow, with the intent of validating the framework developed and provided by Andrés Cremades [6]. The validation consisted of reproducing the findings of the following paper [7] in order to apply the approach to a novel case study. The second simulation involves a smooth channel flow in the presence of temperature, which is treated as a passive scalar.
As previously stated, the analysis is structured in three main steps: firstly, the U-Net is trained on the simulation with the goal of predicting the evolution of 3D snapshots at a temporal distance of 5 time viscous units. Once the model is trained, SHAP values are computed based on the mean-squared error (MSE) between the network prediction and the ground truth, thus computing the importance score of the grid points. Finally, structures are computed from the SHAP values. Different types of conditions for detection of the structures are utilized and compared, with the aim of assessing which criteria lead to the identification to informative structures that can provide new insight from a purely data-driven approach.

2.1. Geometry

The geometry of the problem is a channel, as shown in Fig. 1. Periodic boundary conditions are enforced in the streamwise and spanwise directions. A no-slip condition is enforced at the walls. The domain sizes $L_x \times L_y \times L_z$, mesh sizes $N_x \times N_y \times N_z$, and main parameters of the two simulations performed in this work are summarized in Table 1. For both simulations, the grid spacing in the streamwise and spanwise direction is uniform, while in the wall-normal direction it is governed by a tanh-based grid stretching function of the form

$$y(i) = \frac{1}{2}\left[1 + \frac{\tanh\left(\left(y_0(i) - \frac{1}{2}\right)\alpha_s\right)}{\tanh(\frac{\alpha_s}{2})}\right] \tag{1}$$

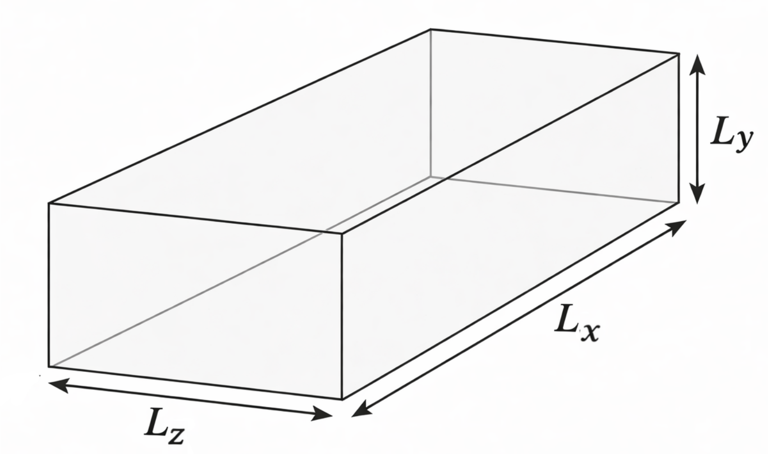


Figure 1: Geometry of the channel flow.

where α_s is the stretching factor, which was taken as 4, and y is the cell-center coordinate in the wall-normal direction. This spacing ensures a fine enough grid resolution in the vicinity of the wall to fully resolve the smaller scales of the turbulent flow, and a coarser one near the center of the channel where less resolution is needed. The mesh and domain size of the first simulation were chosen in order to match the parameters of the simulation performed by Cremades [6], with the goal of validating the framework.
For simulation 2, the Re_τ was increased to 200 with the aim of ensuring that turbulence is sustained and statistically stationary, and with a better separation of scales, while also keeping the computational cost limited to the hardware that was available for this project.

2.2. Governing equations

The turbulent channel flow is considered incompressible, governed by the Navier-Stokes equations and the transport equation for a passive scalar. In the following equations, the variables are normalized by the friction velocity u_τ and the half-channel height δ. The governing equations are:

$$\begin{cases} \dfrac{\partial u_i}{\partial x_i} = 0 \\ \dfrac{\partial u_j}{\partial t} + u_i \dfrac{\partial u_j}{\partial x_i} = -\dfrac{\partial p}{\partial x_j} + \dfrac{1}{Re_\tau} \dfrac{\partial^2 u_j}{\partial x_i^2} - \Pi_j \\ \dfrac{\partial T}{\partial t} + u_i \dfrac{\partial T}{\partial x_i} = \dfrac{1}{Pe_\tau} \dfrac{\partial^2 T}{\partial x_i^2} + Q \end{cases} \tag{2}$$

where u_i are the velocity components, p is the pressure, T is the temperature, Π_j is the constant pressure term that drives the flow through the channel. Regarding the energy equation, it is solved independently of the continuity and momentum equations since they are not coupled. This means that the temperature in the setup is treated as a passive scalar; hence, no buoyancy term appears in the momentum equation, the density is constant and set to unity. The source term Q represents a uniform volumetric heat source applied throughout the domain: this is clearly artificial, but it allows for direct comparison between the structures arising from the momentum and heat transfer by mimicking the effect of the pressure gradient driving the flow. The nondimensional numbers that appear in the governing equations are defined as follows:

$$Re_\tau = \frac{u_\tau \delta}{\nu}, \quad Pr = \frac{\nu}{\alpha}, \quad Pe_\tau = Pr Re_\tau \tag{3}$$

Re_τ is the friction Reynolds number, u_τ is the friction velocity, δ is the half-channel height, ν is the kinematic viscosity, Pr is the Prandtl number, α is the thermal diffusivity, and Pe_τ is the friction Péclet number.

Simulation	$L_x \times L_y \times L_z$	$N_x \times N_y \times N_z$	Re_τ	Pr
1	$2\pi\delta \times 2\delta \times \pi\delta$	$192 \times 201 \times 96$	125	-
2	$2\pi\delta \times 2\delta \times \pi\delta$	$160 \times 200 \times 160$	200	0.5

Table 1: Main parameters of the two simulations, where δ is equal to 1.

2.3. Numerical solver

To perform the direct numerical simulations required in this study, an in-house version of the open-source solver CaNS (Canonical Navier–Stokes), developed by Pedro Costa et al.[5], was employed. The modified version, which differs from the publicly available solver, allows for the computation of channel flow in the presence of multiple passive scalars. The solver combines classical finite-difference approximations with highly optimized FFT-based Poisson solvers, allowing it to achieve parallel scalability on modern CPU and GPU clusters. An overview of the numerical methods implemented in CaNS is proposed here.

The solver employs a structured Cartesian grid and uses a staggered MAC (Marker-And-Cell) layout: velocity components are stored at cell faces, while pressure is stored at cell centers. This classical arrangement improves numerical stability and avoids pressure–velocity decoupling. Spatial derivatives are approximated using second-order central finite differences. Although higher-order schemes exist, the combination of second-order accuracy and uniform Cartesian grids provides a good balance between numerical accuracy and implementation simplicity. Time advancement is performed using a three-stage low-storage Runge–Kutta method. This scheme is explicit, easy to implement, and sufficiently stable for the high-resolution DNS setups for which CaNS is intended. To enforce the incompressibility constraint, pressure projection is needed. Although the Navier–Stokes equations themselves do not include an explicit evolution equation for pressure, the pressure must still adjust instantaneously to ensure that the velocity field remains divergence-free at every time step. CaNS accomplishes this through a classical fractional-step method, with an implementation carefully optimized for structured grids and FFT-based Poisson solvers.

CaNS distinguishes itself through a highly optimized direct Poisson solver based on Fast Fourier Transforms (FFTs). Since the streamwise and spanwise directions are periodic, the Laplacian operator can be diagonalized in the Fourier space. This reduces the 3D problem into a series of independent 1D tridiagonal systems in the wall-normal direction, which are solved with $O(N \log N)$ complexity. This approach allows for massive parallel scalability on modern GPU clusters.

2.4. Deep Learning Architecture for Turbulence Prediction

In recent years, machine learning has emerged as a complementary paradigm for modeling and analyzing turbulent flows, enabling data-driven discovery and reduced-order representations [2, 41]. Among the various data-driven approaches proposed for turbulence prediction, convolutional neural networks have proven particularly effective at learning multiscale spatial correlations in high-dimensional flow fields. In this work, a three-dimensional U-Net architecture is employed as a non-linear surrogate model for short-time turbulent-flow prediction. The network serves as a mapping $\mathcal{F} : x_i'(t) \to x_i'(t + \Delta t)$, where the goal is to evolve the state of the turbulent fluctuations across a discrete temporal step. This section details the architecture, the preprocessing pipeline, and the physical rationale behind the selection of a multiscale convolutional framework, following the methodology established by Cremades et al. [7]. The network performs supervised learning on DNS data of a turbulent channel flow. Input samples consist of the instantaneous fluctuating velocity field $u'(x, y, z, t)$ and, when present, the instantaneous fluctuating temperature field $T'(x, y, z, t)$, expressed on a structured 3D grid. The learning target is the corresponding field evaluated at a fixed temporal offset, chosen to be $\Delta t^+ = 5$ viscous time units. This prediction horizon is strategically kept short relative to the characteristic eddy turnover time; as noted by Cremades et al. [7], this allows the network to focus on the local, nonlinear evolution of coherent structures, before the chaotic nature of the turbulence leads to a complete loss of predictability.

Each DNS snapshot is arranged as a four-dimensional tensor $\boldsymbol{X} \in R^{N_x \times N_y \times N_z \times C}$, where $C \in \{3, 4\}$ represents the input channels. To ensure that the model learns the underlying dynamics rather than the mean flow statistics, the velocity and temperature fluctuations are obtained by subtracting the Reynolds-averaged profile $\overline{u}(y)$ and $\overline{T}(y)$ respectively. These fluctuations are then normalized through a min-max scaling, which facilitates stable gradient propagation.

Because the DNS domain is periodic in the streamwise and spanwise directions, periodic padding is applied to every input sample prior to the convolutional layers. In standard image processing, zero-padding is often used to maintain spatial dimensions; however, in fluid dynamics, this introduces non-physical "walls" at the boundaries that disrupt the continuity of the flow. By copying grid values from one boundary to the opposite side, the 3D kernels operate on a topologically continuous domain, preserving the translational invariance and the physical symmetries of the periodic channel.

2.4.1 U-Net architecture

U-Net is a convolutional neural network architecture originally introduced for biomedical image segmentation [33], where accurate localization and multi-scale feature recovery are essential for reconstructing fine structural details. Its encoder–decoder structure with skip connections has since been adopted in computational physics, where similar requirements arise in the reconstruction and prediction of spatially complex flow fields. In fluid dynamics applications, U-Net–based models have been successfully used to infer velocity and pressure distributions in turbulent flows within Reynolds-averaged Navier–Stokes frameworks [40] and to reconstruct high-resolution turbulent fields from coarse or downsampled data [11]. These studies demonstrate that the multi-scale feature hierarchy and resolution-preserving skip connections of U-Net architectures are particularly well suited for representing the wide range of spatial scales characteristic of turbulent flows. Motivated by these properties, the choice of a 3D U-Net architecture is particularly suitable to learn mappings between flow fields while preserving both large-scale structures and small-scale turbulent features. In conventional convolutional neural networks, the encoder (or contracting path) progressively extracts higher-level feature representations by combining convolution and downsampling operations. While this process increases the receptive field and enables the network to capture global contextual information, it also reduces spatial resolution and can obscure precise localization of fine-scale structures. For image-to-image regression tasks such as flow-field reconstruction, this loss of spatial detail poses a significant limitation, since accurate prediction requires preserving the geometric coherence of small-scale features across the domain. The U-Net architecture addresses this issue by coupling the encoder with a symmetric decoder that restores spatial resolution through learned upsampling, while skip connections directly transfer high-resolution features from early layers to later stages of the network [33]. This design allows the model to integrate global contextual information with localized detail, which is essential in turbulent flows. For the present study, the U-Net adopts a symmetric 3D encoder–decoder configuration designed to process volumetric flow data and capture correlations across all spatial directions. The encoder consists of multiple resolution levels in which each block applies repeated 3D convolutions with $3 \times 3 \times 3$ kernels, followed by batch normalization and Rectified Linear Unit (ReLU) activation. Spatial downsampling is performed using 3D max-pooling, which progressively enlarges the receptive field while reducing spatial resolution. The number of feature channels increases with network depth, enabling the model to represent increasingly abstract flow features at coarser scales. At the bottleneck, the compressed representation aggregates global spatial interactions and encodes large-scale flow structures that influence the evolution of smaller turbulent motions. The decoder mirrors the encoder architecture and restores spatial resolution using 3D transposed convolutions. At each level, skip connections concatenate encoder feature maps with the upsampled decoder activations, ensuring that high-resolution information is preserved and reintegrated during reconstruction.

The skip-connection mechanism plays a central role in maintaining the fidelity of reconstructed flow fields. During the encoding process, repeated pooling operations inevitably attenuate high-frequency spatial content associated with sharp gradients and small-scale vortical structures. By directly concatenating encoder feature maps with decoder activations at matching resolutions, the network reintroduces fine-grained spatial information that would otherwise be difficult to recover from the compressed bottleneck representation alone. In the context of turbulent flow reconstruction, this mechanism can be interpreted as enabling the simultaneous propagation of large-scale contextual features and localized structures, allowing the model to represent interactions across multiple spatial scales. As a result, the decoder is guided not only by global latent features but also by high-resolution cues that help preserve the geometric consistency of near-wall gradients and small-scale turbulent patterns in the predicted fields.

Here, a brief description of the mathematical operations that are present in the U-Net architecture is proposed: The non-linear transformation within each convolutional block is governed by the Rectified Linear Unit (ReLU) activation function. ReLU introduces the non-linearity necessary to model the convective terms and complex modal interactions of turbulent flows. For any input scalar z within the feature map tensor after batch

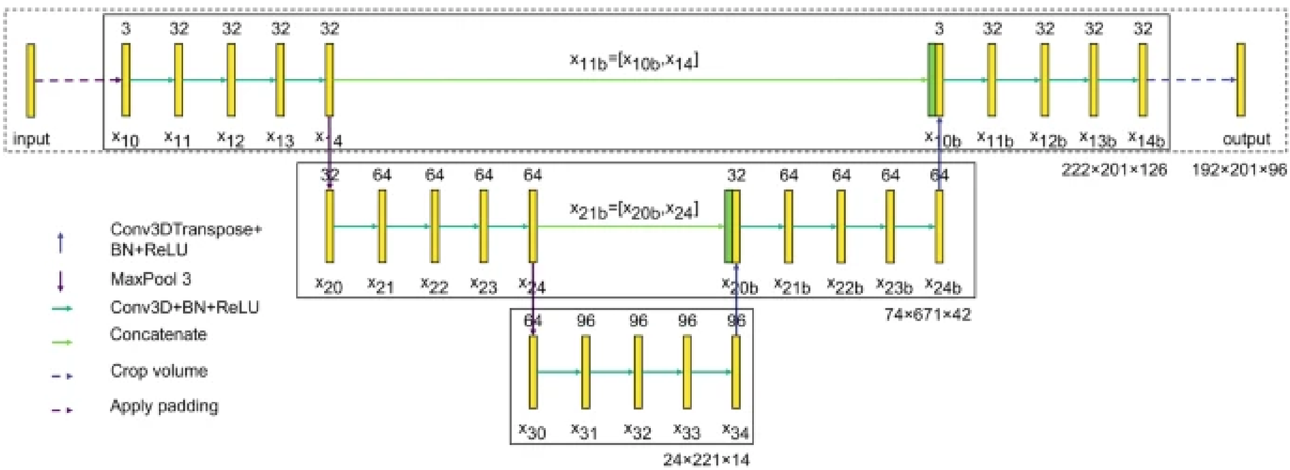


Figure 2: Schematics of the U-Net architecture from Cremades et al. [6].

normalization, the ReLU operation is defined as:

$$\text{ReLU}(z) = \max(0, z) \tag{4}$$

Applied element-wise to the pre-activation tensor $\mathbf{Z} \in \mathbb{R}^{C\times D\times H\times W}$, the mapping ensures that all negative activations are suppressed to zero while positive values are passed through unchanged.

$$(x_{i,k})_{d,h,w} = \text{ReLU}((z_{i,k})_{d,h,w}) = \begin{cases} (z_{i,k})_{d,h,w} & \text{if } (z_{i,k})_{d,h,w} > 0 \\ 0 & \text{if } (z_{i,k})_{d,h,w} \leq 0 \end{cases} \tag{5}$$

where $z_{i,k}$ represents the k-th feature map of layer i after the 3D convolution and batch normalization operations. The primary building block of the network is the 3D Convolutional layer, which extracts spatial correlations across the streamwise, wall-normal, and spanwise directions. The transformation for the k-th feature map at layer i is given by:

$$x_{i,k} = \text{ReLU}\left(\gamma_k\left(\frac{(W_k * \mathbf{X}) + b_k - \mu_k}{\sqrt{\sigma_k^2 + \epsilon}}\right) + \beta_k\right) \tag{6}$$

where $x_{i,k}$ is the resulting k-th feature map of layer i, representing the output activation volume after non-linearity, $\mathbf{X}$ is the input tensor to the layer, b_k is the learnable bias parameter associated with the k-th filter, ϵ is a small positive constant, μ_k and σ_k^2 are respectively the mean value and the variance of the k-th feature map calculated over the mini-batch during the Batch Normalization step, and lastly γ_k and β_k are the learnable scale and shift parameters that allow the network to linearly transform the normalized activations.
The 3D cross-correlation operation evaluated at the spatial coordinate (d, h, w), which aggregates local spatial and channel-wise information, is given by

$$(W_k * \mathbf{X})_{d,h,w} = \sum_{c=1}^{C_{in}} \sum_{i=0}^{k_d-1} \sum_{j=0}^{k_h-1} \sum_{l=0}^{k_w-1} W_{k,c,i,j,l} \cdot \mathbf{X}_{c,d+i,h+j,w+l} \tag{7}$$

$W_{k,c,i,j,l}$ are the learnable weights of the k-th convolutional kernel for the c-th input channel at the local filter coordinates (i, j, l), k_d, k_h, k_w are the kernel dimensions for depth, height, and width, respectively. Spatial downsampling is achieved via 3D Max Pooling. This operation reduces the degrees of freedom while increasing the receptive field, allowing the network to capture large-scale energy-containing eddies at deeper layers. For an input tensor x_L, the value at coordinate (d, h, w) in the subsequent level x_{L+1} is defined as:

$$x_{L+1}(d, h, w) = \max_{0\leq \delta d,\delta h,\delta w<\kappa} x_L(s \cdot d + \delta d, s \cdot h + \delta h, s \cdot w + \delta w) \tag{8}$$

where s is the stride, representing the step size the pooling window or kernel moves across the input volume, κ is the pooling kernel size, defining the dimensions of the 3D window $(\kappa \times \kappa \times \kappa)$ from which the maximum value is extracted, $\delta d, \delta h, \delta w$ are the local offset indices within the pooling window, ranging from 0 up to $\kappa - 1$.
To reconstruct the prediction at the original DNS resolution, the decoder utilizes 3D Transposed Convolutions [8]. This operation projects low-resolution latent representations y to a high-resolution output z:

$$z_{d,h,w} = \sum_{i,j,l} K_{i,j,l} \cdot y_{\frac{d-i}{s},\frac{h-j}{s},\frac{w-l}{s}} \tag{9}$$

$K_{i,j,l}$ are the learnable weights of the 3D transposed convolution kernel at the local filter coordinates (i, j, l), $\frac{d-i}{s}, \frac{h-j}{s}, \frac{w-l}{s}$ are the fractional indices used in the transposed convolution; in practice, these correspond to the coordinates in the low-resolution input y that contribute to the high-resolution output z. Unlike simple interpolation, transposed convolution allows the network to learn the optimal way to upsample the flow field, ensuring that the reconstructed small-scale patterns are physically consistent with the large-scale context provided by the bottleneck. This upsampling is further refined by the skip-connection mechanism, which concatenates high-resolution encoder features to compensate for the information loss incurred during the max-pooling stages.

2.4.2 Training Procedure

The network is trained using Mean Squared Error (MSE) loss function:

$$\mathcal{L} = ||\hat{\boldsymbol{u}}(t + \Delta t) - \boldsymbol{u}(t + \Delta t)||_2^2 \tag{10}$$

where $\hat{\boldsymbol{u}}(t + \Delta t)$ is the predicted velocity field and $\boldsymbol{u}(t + \Delta t)$ is the ground-truth. This loss is appropriate for continuous regression tasks and provides a direct metric of pointwise prediction accuracy. The optimiser used

Simulation	Fields	η	Filters	Epochs	ρ	Channels
1	4000	1×10^{-5}	32	400	0.9	3
2	4000	1×10^{-5}	32	240	0.9	4

Table 2: Main parameters of the training for the three simulations. In particular, ρ is the decay rate, η is the learning rate; these parameters appear, respectively, in Equation 12 and Equation 13.

is the RMSprop (Root Mean Square propagation), proposed by [13]: it is an adaptive learning-rate method for stochastic gradient descent. Unlike standard Stochastic Gradient Descent (SGD), which uses a global learning rate, RMSprop rescales the update for each parameter based on a running average of the magnitude of recent gradients. The optimization process follows a three-step iterative cycle; the structure of the algorithm is as follows: in the first step, the parameters are initialized and the gradient of the loss function with respect to the parameters at step t is computed as

$$g_t = \nabla_\theta \mathcal{L}(\theta_t) \tag{11}$$

where θ_t denotes the vector of all trainable parameters at the optimisation step t. Then, in the second step, the moving average of the squared gradients $E[g^2]$ is updated

$$E[g^2]_t = \rho E[g^2]_{t-1} + (1-\rho) g_t^2 \tag{12}$$

where ρ is the decay factor for the moving average, usually set to 0.9. Finally, in the third step, the vector of all the parameters is then updated for $t+1$ using learning rate η, adjusted using the moving average of the squared gradients,

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \varepsilon}} g_t. \tag{13}$$

where ε is a small constant added to prevent division by zero.
Concerning the partitioning, the datasets were formed by 6000 instantaneous fields, divided into training, validation, and test subsets. Out of the available fields, 4000 were allocated to training and validation. 80% of the 4000 were used for training and the remaining 20% was used for validation. During the training phase, the training dataset is seen by the network in full for every epoch, but the order in which the fields were seen in every epoch was randomized. The shuffling is essential for fulfilling the independent and identically distributed assumption underlying the gradient descent framework, ultimately allowing the 3D U-Net to learn the universal spatial correlations of the turbulent flow rather than specific temporal trajectories. The remaining 2000 fields of the data was not seen during training and formed an independent test set for final performance evaluation and for the computation of the SHAP values. The training of the models was performed using a NVIDIA RTX5090 GPU. The significant memory capacity and computational power of this hardware were instrumental in handling the volumetric DNS datasets and the high-dimensional Gradient-SHAP calculations. The main parameters for the training for the three cases are summarized in Table 2. As it can be seen, most of the parameters have been kept the same; in particular, even when a fourth channel (the temperature field) is added, the architecture of the U-Net remains virtually unchanged: only the first layer in the encoding path and the last layer in the decoding path were adjusted by adding one filter (from 3 to 4) in order to keep the dimensions consistent.

2.5. SHAP values

Explainable artificial intelligence has emerged as a critical framework for interpreting complex machine-learning models and extracting physically meaningful insight from high-dimensional predictions [28]. Attribution-based interpretability methods aim to assign relevance scores to input features according to their contribution to a model's output [37]. Among attribution-based methods, SHAP (SHapley Additive exPlanations), introduced by Lundberg and Lee [27], provides a principled approach for quantifying the contribution of individual input features. SHAP is grounded in cooperative game theory, specifically adapting the concept of Shapley value, a solution concept introduced by Lloyd Shapley in 1953 [36]. Originally developed to determine the fair distribution of a total payout among players in a coalitional game, the Shapley value represents the weighted average of a player's marginal contribution across all possible permutations of the group. This approach ensures that the distribution satisfies four fundamental axioms of fairness: symmetry, dummy, additivity, and efficiency [4]. In the context of machine learning, SHAP provides a mathematically rigorous framework for attribution-based feature importance by treating the prediction task as a coalitional game. Here, the "players" are the input features, and the "payout" to be distributed is the difference between the model's prediction for a specific instance and the expected prediction across the dataset. The axiomatic guarantees ensure mathematical consistency of the attribution, although physical interpretability depends on the predictive fidelity of the surrogate model. The axiom of efficiency dictates that the total importance assigned to the input field must sum exactly to the difference between the model's prediction and its expected base value, ensuring that the model-attributed contribution is quantitatively tied to the magnitude of the predicted fluctuation [28]. The symmetry axiom ensures that input features that contribute identically to the prediction receive identical attribution, which is consistent with the statistical symmetries of the channel flow when these are represented in the data. Furthermore, the dummy axiom assigns zero attribution to features that do not influence the model output, thereby helping distinguish dynamically inactive regions. Finally, the additivity property permits consistent attribution across linear combinations of models, facilitating comparative analysis of the contributions associated with different physical variables.

While these properties define the theoretical foundation of SHAP, the exact computation of Shapley values becomes computationally intractable for high-dimensional inputs such as three-dimensional turbulent flow fields, due to the combinatorial growth in feature subsets. Consequently, practical implementations rely on efficient approximations that retain the core axiomatic structure while remaining scalable for deep neural networks.

In this work, the gradient-SHAP algorithm is applied to compute the importance scores of individual grid points in predicting the output velocity fields. Unlike traditional SHAP, which can be computationally prohibitive for 3D volumetric data, Gradient-SHAP utilizes an Expected Gradients (EG) approach [9]. To assess the accuracy of the flow predictions, an additional layer is appended to the model output. This layer evaluates the mean-squared error (MSE) between the predicted flow and the ground-truth field at the subsequent time step. As a result, the original output of the model is reduced to a single scalar value representing the overall prediction error. The contribution of each grid point to the MSE is measured through its local SHAP value, expressed as

$$\mathrm{MSE} = \phi_0 + \sum_{j\in\{u,v,w,T\}} \sum_{i=0}^{N} \phi_{ji}\, z_{ji}, \tag{14}$$

where ϕ_0 denotes the MSE associated with the reference (non-informative) field, ϕ_{ji} is the SHAP value corresponding to velocity component j at grid point i, and z_{ji} is a binary indicator that equals 0 when the grid point value is taken from the reference field and 1 when it is taken from the input field.

In this work, the standard EG formulation is slightly modified to evaluate feature importance. While the original EG computes SHAP values as expectations over a set of reference fields, the present approach uses a single reference field, defined as the mean velocity field. The method is described by

$$\phi_i(x_{\mathrm{in}}) = \mathbb{E}_{x_{\mathrm{ref}},\, \alpha\sim U(0,1)} \left[(x_{\mathrm{in},i} - x_{\mathrm{ref},i}) \frac{\partial F\big(x_{\mathrm{ref}} + \alpha(x_{\mathrm{in}} - x_{\mathrm{ref}})\big)}{\partial x_{\mathrm{in},i}} \right], \tag{15}$$

where $\phi_i(x_{\mathrm{in}})$ represents the SHAP value of feature i in the input field x_{in}, x_{ref} is the reference field, $\mathcal{F}$ denotes the model that computes the MSE of the predictions, and $\alpha \in [0, 1]$ interpolates between the reference and input fields. Computing the expected gradients for all grid points yields a spatial field that reflects the relative importance of each input node in the prediction. Because the gradient-based computation of SHAP values can introduce high-frequency numerical noise, a smoothing strategy leveraging the physics of the flow is employed. We exploit the periodic boundary conditions of the turbulent channel by performing streamwise and spanwise shifts. Three separate importance fields are computed by displacing the input field along its periodic directions. These physically equivalent realizations are then averaged

$$\bar{\phi}(x, y, z) = \frac{1}{N_{\mathrm{shift}}} \sum_{s=1}^{N_{\mathrm{shift}}} \phi_s(x, y, z). \tag{16}$$

This procedure suppresses stochastic noise while highlighting the persistent, large-scale coherent structures.

2.6. Classically-studied Coherent Structures

Coherent structures in turbulent flows are spatially and temporally organized motions that make a disproportionate contribution to transport, production, or energy relative to the surrounding fluctuations. Although turbulence arises from deterministic governing equations, nonlinear multiscale interactions produce velocity fields that appear irregular; embedded within them, however, are dynamically significant patterns whose identification depends on the diagnostic framework adopted. In Eulerian analyses, coherent vortical motions are frequently extracted from invariants of the velocity gradient tensor. The Q-criterion identifies regions where rotation locally exceeds strain [14], while the λ_2 criterion refines vortex-core detection through eigenvalue analysis of combined strain–rotation tensors [15]. These approaches provide objective, instantaneous markers of topology and have been widely used in direct numerical simulations of wall turbulence. Their limitation lies in the necessity of threshold selection and in their emphasis on kinematic structure rather than direct transport significance. A complementary Lagrangian perspective identifies coherent structures as material surfaces that organize transport over finite times. Finite-time Lyapunov exponent (FTLE) fields reveal attracting and repelling transport barriers that shape scalar mixing and momentum redistribution [35]. Such Lagrangian coherent structures provide a dynamically meaningful partition of the flow but require time-resolved velocity data and substantial computational effort. Modal decomposition techniques offer yet another viewpoint. Proper orthogonal decomposition (POD) extracts energetically optimal modes from turbulent datasets [26], and has been extensively applied to wall-bounded flows to reveal dominant large-scale and very-large-scale motions [1]. While powerful for reduced-order modeling, modal approaches optimize mathematical criteria such as energy capture and may mix distinct physical mechanisms within individual modes, complicating interpretation in terms of momentum transport.

In wall-bounded turbulence, the defining dynamical quantity is the Reynolds shear stress, which mediates momentum transfer from the outer flow toward the wall. Early experimental studies established that this transport is highly intermittent and dominated by burst-like events in the near-wall region [20]. Quadrant analysis provides a statistical framework for isolating these events by classifying instantaneous velocity fluctuations according to the signs of the streamwise and wall-normal components [43]. Of the four resulting categories, ejections (Q2 events) and sweeps (Q4 events) contribute most strongly to the Reynolds shear stress and to turbulent kinetic energy production. Ejections correspond to low-speed fluid lifted away from the wall, while sweeps represent high-speed fluid impinging toward it. The introduction of a threshold, or hole size, isolates intense events and demonstrates that a relatively small fraction of strong Q2 and Q4 events accounts for a large share of the total momentum transport [25].

Compared with purely topological or modal methods, Q event analysis is directly tied to the physics of wall shear stress and energy production. It requires only pointwise velocity measurements and is therefore well suited for experiments, as well as numerical simulations. Although it does not explicitly reconstruct three-dimensional geometry and depends on threshold selection, it provides a physically transparent bridge between intermittent bursting dynamics and sustained momentum transfer. For this reason, quadrant-based identification remains one of the most robust and interpretable tools for analyzing coherent structure dynamics in turbulent shear flows.

In flows with heat transfer, the quadrant framework can be extended to thermal Q events, which classify combined fluctuations of wall-normal velocity and temperature fields [3, 34]. Thermal Q events separate motions according to the relative signs of wall-normal velocity and temperature fluctuations. Upward-moving warm fluid corresponds to $Q1_T$, while upward-moving cool fluid, representing thermal ejections, is $Q2_T$. Downward-moving cool fluid is $Q3_T$, and downward-moving warm fluid, representing thermal sweeps, is $Q4_T$. Intense thermal Q2 and Q4 events dominate convective heat transport, analogous to the way velocity-based Q2 and Q4 events dominate Reynolds shear stress. When temperature behaves as a passive scalar, its influence on coherent structures is fundamentally one-way: the velocity field organizes the scalar field, but the scalar does not modify the underlying momentum dynamics. In this regime, thermal Q events do not represent dynamically self-sustaining structures, but rather scalar signatures of velocity-driven motions such as ejections and sweeps. The spatial organization, frequency, and intensity of $Q2_T$ and $Q4_T$ events therefore closely mirror those of their momentum counterparts, with differences arising primarily from scalar diffusivity and molecular Prandtl number effects. Passive transport leads to sharper scalar gradients and enhanced intermittency at high Prandtl number, while lower Prandtl numbers promote smoother scalar fields and weaker small-scale clustering. In the context of this work, the only classical coherent structures considered are Q events and thermal Q events, which are then compared with structures segmented through the SHAP methodology. The Q events are defined as connected regions satisfying

$$|u'(x, y, z, t)\, v'(x, y, z, t)| > \beta\, u'_{rms}(y)\, v'_{rms}(y) \tag{17}$$

as suggested by Lozano-Durán et al. [24]. $|u'(x, y, z, t)v'(x, y, z, t)|$ represents the instantaneous point-wise tangential Reynolds stress and β is the percolation index specific for Q events. Throughout this study, the subscript rms is used to represent the root mean square of the respective quantities. Analogously, the thermal Q events are defined as connected regions satisfying

$$|v'(x, y, z, t)\, T'(x, y, z, t)| > \beta_T\, v'_{rms}(y)\, T'_{rms}(y) \tag{18}$$

where $|v'(x, y, z, t)\, T'(x, y, z, t)|$ is the turbulent heat flux and β_T is the percolation index specific for thermal Q events.

It should be clarified that many identification methods for Q events have been proposed. For instance, in the original analysis of these types of structures, the thresholding condition was based on the local mean Reynolds stress [45], while more recent works suggested the use of the root mean square of the stress fluctuations [31]. Different conditions will inevitably lead to different values of the percolation indices. Furthermore, some conditions may be ill-suited in certain types of flows. For instance, using identification methods with a constant threshold, with some reference y_0, is problematic in inhomogeneous flows due to the sensitivity of the results with respect to the reference values [29]. Furthermore, it is important to emphasize that quadrant structures are fundamentally kinematic in nature. They are defined solely through the instantaneous signs and amplitudes of the velocity fluctuations, without reference to spatial connectivity, vorticity topology, or dynamical invariants. As such, they do not constitute coherent structures in a geometric sense, but rather statistical events associated with specific contributions to the Reynolds shear stress $-u'v'$. Their definition is local and instantaneous. This kinematic character is both a limitation and a strength. On one hand, quadrant analysis does not reveal the full spatial organization of vortical packets, streak instabilities, or large-scale motions. On the other hand, precisely because it avoids geometric assumptions, it provides a direct measure of the instantaneous transport mechanisms responsible for shear stress production. In this sense, quadrant events isolate the transport-active states of the flow, independently of the specific structural morphology that generates them.

2.7. Identification and Characterization of SHAP-based Structures

After applying the gradient-SHAP framework to the problem, a vector field is obtained that characterizes the relevance of each spatial region to the flow dynamics. This field is given by the SHAP values $\boldsymbol{\phi}$, which quantify the causal contribution of each individual grid point to the temporal evolution of the flow. As discussed above, the SHAP values indicate which locations in the domain exert the greatest influence on the model evolution. Consequently, the resulting SHAP-based structures highlight the flow regions that should be targeted for effective flow control.

The SHAP vector provides separate importance measures for the streamwise, wall-normal, and spanwise velocity components and temperature when present, denoted by ϕ_u, ϕ_v, ϕ_w, and ϕ_T, respectively.

The segmentation of discrete SHAP structures from the continuous importance field $\boldsymbol{\phi}$ is formalized through a multidimensional percolation process. This approach is grounded in the necessity of isolating dynamically significant "events" from background attribution noise. By systematically varying a threshold parameter H, the field is decomposed into a maximal number of independent, disjointed volumes.

This provides a topology-based thresholding strategy, although it still depends on the chosen normalization and percolation parameter. Rather than relying on arbitrary magnitudes, the process exploits the topology of the attribution field to identify the cores of the structures. Consequently, the resulting structures represent the primary spatial domains where the neural network identifies a deterministic link between the input state and the future evolution of the flow, providing a robust basis for the subsequent causal importance analysis.

The criterion for the detection of the structures used in this work reads

$$\sqrt{\phi_u^2(x,y,z,t)+\phi_v^2(x,y,z,t)+\phi_w^2(x,y,z,t)+\phi_T^2(x,y,z,t)} > H\sqrt{\overline{\phi_u^2}(y)+\overline{\phi_v^2}(y)+\overline{\phi_w^2}(y)+\overline{\phi_T^2}(y)}, \tag{19}$$

where H is the percolation parameter. To extract discrete, physically meaningful structures from the SHAP importance field, a thresholding operation is required. In this work, H is determined through a sensitivity analysis aimed at maximizing the total number of individual structures within the computational domain. In particular, if H is set too low, the identified regions tend to merge into a single "sponge-like" structure that spans the entire domain, providing little insight into local dynamics. Conversely, if H is set too high, only the most extreme peaks of importance are captured, leading to a sparse representation that may neglect significant, albeit less intense, causal interactions. By identifying the value of H that yields the highest count of disjointed structures, we ensure that the threshold effectively separates the "background" importance from the primary coherent events, capturing the granularity of the turbulent features. This optimization is performed independently for each flow.

Traditional methods identify coherent structures using kinematic or dynamic criteria, such as the aforementioned tangential Reynolds stress in Q events. While these methods are physically grounded, they are limited to the specific criteria defined by the researcher. SHAP-based structures are identified through a purely data-driven paradigm that evaluates importance based on the inputs predicictive impact. This allows for the discovery of influential regions that may not align with traditional definitions, such as regions where velocity-temperature coupling is high but instantaneous momentum transport is low. By classifying regions based on their contribution to the evolution of the field, SHAP could prove to be an effective auxiliary framework when paired with classical methods for the segmentation of coherent structures described earlier.

For the first simulation, the SHAP vector is only composed of the three components $\boldsymbol{\phi} = (\phi_u, \phi_v, \phi_w)$. These SHAP-based structures are then compared with classical definitions of Q events. While for the second case study, the SHAP vector is composed by $\boldsymbol{\phi} = (\phi_u, \phi_v, \phi_w, \phi_T)$, thus Q events and thermal Q events are compared to different sets of SHAP structures which are the result of different thresholding conditions. Our aim is to identify which condition will yield instructive information regarding the regions of interest of the flow, particularly in a thermal transfer sense.

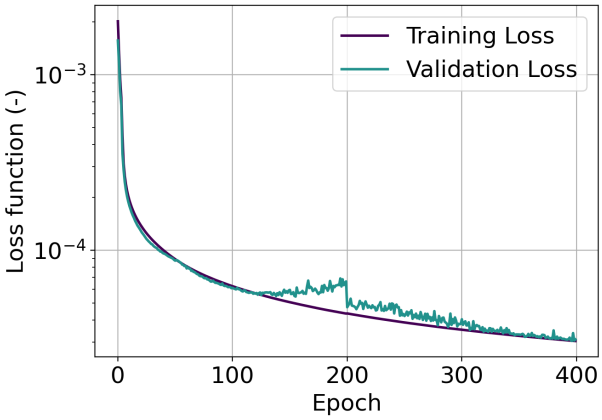


Figure 3: Training and validation loss of the neural network for simulation 1

3. Results

In this section, the main results of the analysis are shown. Firstly, the results of the validation are presented and are compared to the results obtained in the work where the framework was first presented [7]. Subsequently, the results from the channel flow simulations where temperature is also present are illustrated.

3.1. Framework validation

The first analysis was aimed at reproducing the results that obtained by Cremades et al. [7] in which the authors studied a smooth channel flow at a $Re_\tau = 125$. The main geometrical and physical parameters are shown in Table 1. Convergence was ensured by checking both the bulk velocity convergence and the shear stress distribution across the channel. As can be seen in Figure 3, the training consisted of 400 epochs in which every snapshot of the training set is seen by the network. The training loss has a downward trajectory and begins to plateau toward the end. With regard to the validation loss, it also has a downward trajectory that closely follows the training loss. These behaviors are satisfactory, and, to prevent potential overfitting, training was stopped at 400 epochs before a clear plateau was reached. The final values of training and validation loss are respectively $3.043 \cdot 10^{-5}$ and $3.108 \cdot 10^{-5}$, which were judged as adequate.

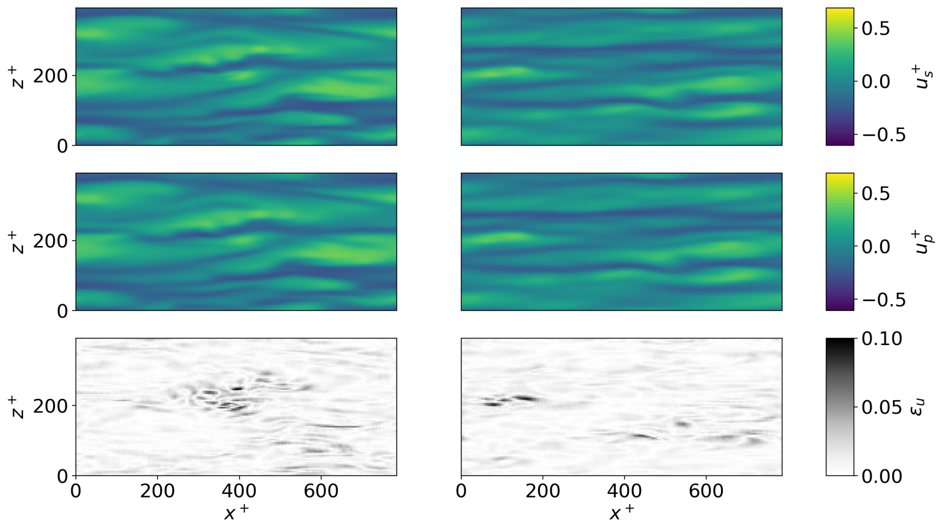
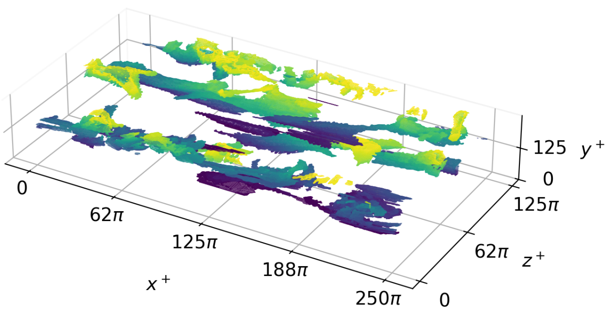
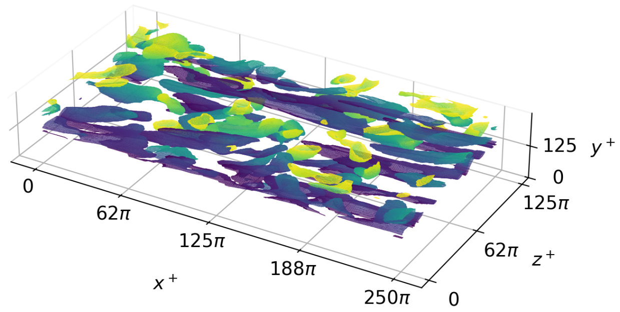


Figure 4: Slice of the $x - z$ plane of the predicted field u_p^+ (first row), the simulated field u_s^+ (second row), and point-wise error between the two (third row) at a wall distance of $y^+ = 10$. The left column is the lower channel slice and the right column is the upper channel slice.



(a)

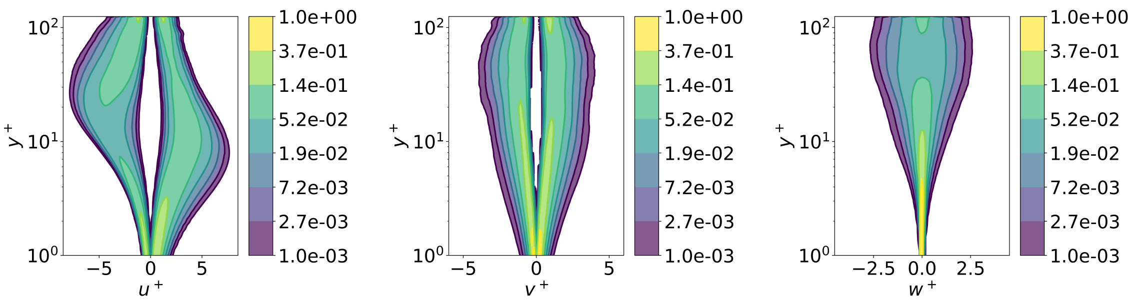


(b)

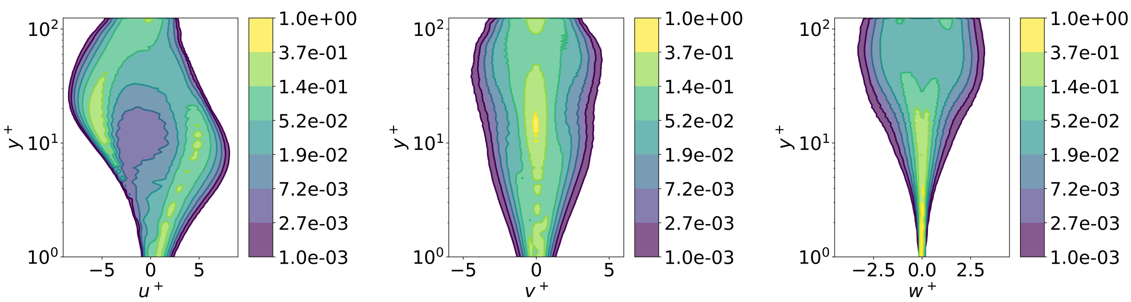
Figure 5: Instantaneous visualization of the coherent structures in the channel flow. In the figure, the wall is located at $y^+ = 0$, and the mid-plane of the channel is $y^+ = 125$. The structures are colored depending on the distance from the wall, from purple (close to the wall) to yellow (close to the centerline). SHAP-based structures are shown in subfigure (a) and Reynolds-stress structures are shown in subfigure (b). The structures shown are the ones in the lower half of the channel.

In Figure 4, a cross-section of an instantaneous snapshot at a distance from the wall of $y^+ = 10$ of the simulated field u_s^+, predicted field u_p^+ and normalized error ε_u between the two is shown as an example. The left column is the lower channel and the right column is the upper channel. It is possible to see that the error never exceeds a point-wise percentage of 10%. The normalized error average was computed for the 2000 fields which were not seen by the network during the training phase, yielding the following results

$$\varepsilon_u = 0.7320\% \quad \varepsilon_v = 0.7678\%, \quad \varepsilon_w = 0.8763\%$$



(a) (b) (c)



(d) (e) (f)

Figure 6: Joint probability density function of the streamwise velocity fluctuation, and the inner-scaled wall-normal distance, y^+ of the structures. The figure presents the SHAP structures, subfigures a), c), and e), Reynolds-stress structures b), d), f).

The point-wise errors were calculated as follows

$$\varepsilon_u = \frac{|u_p^+ - u_s^+|}{\max(u_s^+)}, \quad \varepsilon_v = \frac{|v_p^+ - v_s^+|}{\max(v_s^+)}, \quad \varepsilon_w = \frac{|w_p^+ - w_s^+|}{\max(w_s^+)}, \tag{20}$$

where u, v, w, are respectively the three velocity components in the x, y, z directions. These results are in line with the results from the original paper, where the relative error obtained was within 2%.
After the training, the SHAP values for 2000 fields were computed. As previously stated, from the SHAP values, it is possible to highlight the regions of interest in the flow through the definition of structures using Equation 19 (where ϕ_T was zero in this case). Before computing the actual structure, it was also necessary to identify the value of the percolation parameter H, which was set in order maximize the total number of structures. Once the percolation index was chosen, the SHAP-based structures were identified for every instantaneous field from the corresponding SHAP vector fields $\boldsymbol{\phi}$. An analogous process was carried out to identify the classical structures based on the Reynolds stress, for which the defining condition is the one reported in Equation 17.
In order to check the consistency of the structures computed both through the SHAP values and through the Reynolds stresses, the joint probability density function (JPDF) of the streamwise velocity fluctuation and the inner-scaled wall-normal distance of the structures was computed. The resulting JPDFs are shown in Figure 6, where a), b), and c) are the plots concerning the Reynolds-stress structures and d), e), and f) are the ones concerning the SHAP-based structures. These results were confronted with the findings of Cremades et al. [7]. The general form of the JDPFs of the streamwise component of the velocity is well-captured and consistent with the results of the reference work. Within the viscous sublayer, regions of positive streamwise fluctuations align closely with areas of Q events, which are typically linked to high-velocity sweep structures moving toward the wall. As illustrated in Figure 6, these high-importance, high-velocity regions near the wall remain consistent. Furthermore, another significant area of importance is identified from $y^+ \simeq 30$ toward the channel center; this region corresponds to a high probability of Q events associated with ejections, or low-velocity fluid moving away from the wall.
Although SHAP structures specifically isolate regions of high significance for low-velocity occurrences, Q events enhance the likelihood of sweeps. Consequently, the SHAP framework allows for the extraction of specific Q events that exert a superior influence on flow development, effectively distinguishing the most causal Q events. Nonetheless, for distances $y^+ > 50$, the zones of significant importance where the streamwise velocity fluctuation u is near zero can be attributed to the occurrence of vortices and areas of high dissipation [7]. Even though SHAP structures demonstrate notable similarities to these conventional structures across various sections of the channel, their spatial distributions do not align perfectly with any single traditional category.

3.2. Channel flow with temperature

In this section, we analyze a turbulent smooth channel flow at $Re_\tau = 200$ and $Pr = 0.5$. The main geometrical parameters are reported in Table 1. Compared to the previous configuration, the framework was extended to include temperature as a fourth input channel. This modification did not alter the overall neural network architecture, only the dimensionality of the input–output tensors was increased. Statistical convergence was first verified through bulk flow quantities. Once convergence was achieved, 6000 instantaneous fields were sampled with a temporal spacing of $\Delta t^+ = 5$, consistent with the previous case. The dataset was partitioned identically: 4000 fields were used for training and validation (80% for training and 20% for validation), while the remaining

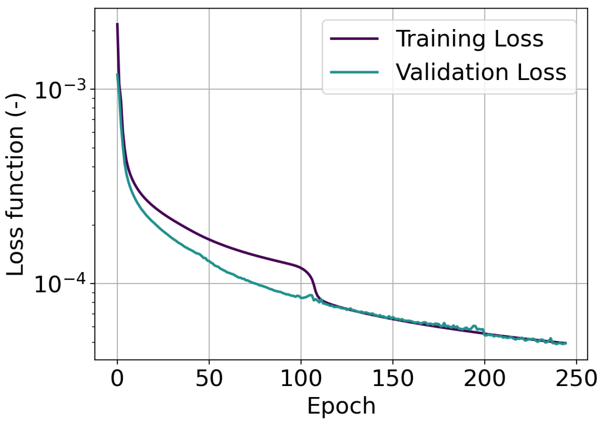
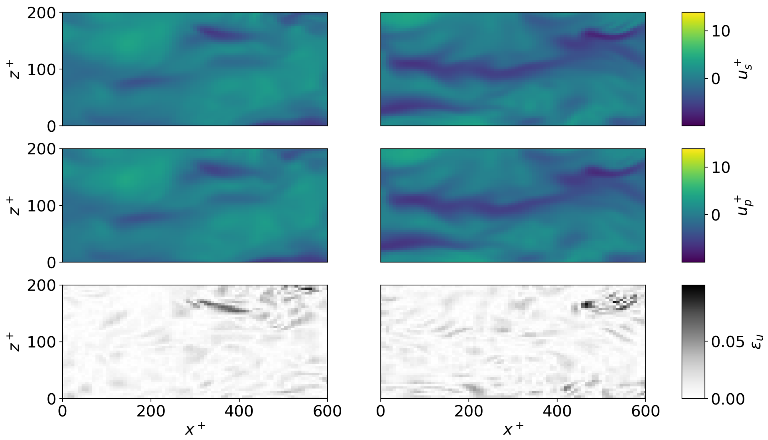
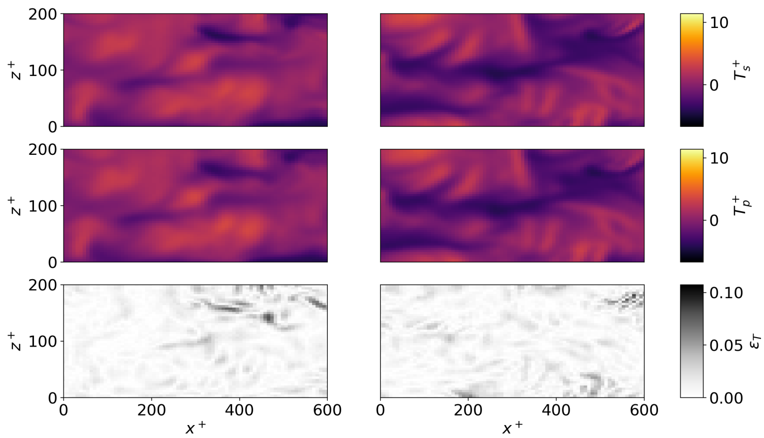


Figure 7: Training and validation loss of the neural network for the second simulation.



(a)



(b)

Figure 8: Slice of the $x - z$ plane of the predicted field (first row), simulated field (second row), and point-wise prediction error (third row) of a) the streamwise velocity component and b) the temperature. The left column is the lower channel slice and the right column is the upper channel slice at a y^+ distance of 28.

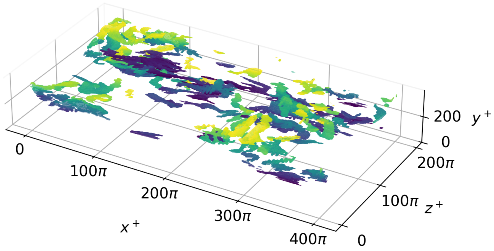
2000 fields were retained as an independent test set for SHAP analysis. The training hyperparameters are summarized in Table 2.
As shown in Figure 7, training was carried out for 240 epochs. During the first 110 epochs, the training loss exceeds the validation loss, indicating an initial optimization imbalance. Beyond this stage, both losses decrease steadily and converge to nearly identical values. Training was stopped at epoch 240, prior to full convergence, due to small but persistent oscillations in the validation loss. Although limited in magnitude, these fluctuations could indicate the onset of overfitting. The final training and validation losses are $4.941 \cdot 10^{-5}$ and $4.951 \cdot 10^{-5}$, respectively. Prediction accuracy was quantified using the normalized mean error over the 2000 test fields that were not used during training, yielding the following results:

$$\varepsilon_u = 0.7282\% \quad \varepsilon_v = 0.8989\%, \quad \varepsilon_w = 0.9046\% \quad \varepsilon_T = 0.7272\%$$

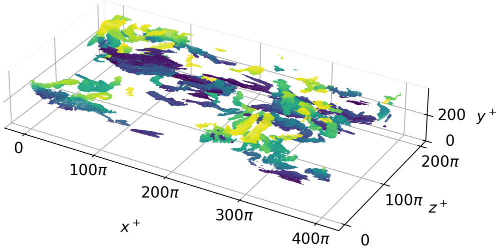
As we can see, the relative averaged error is less than 1%. Note that while the averaged error is extremely low, point-wise error can reach a value of over 10%, as shown in Figure 8, where a slice of the prediction is shown for the streamwise component of the velocity 8a and temperature 8b. Unlike the previous case, two separate SHAP fields were computed for each of the 2000 snapshots, for both velocity and temperature. The difference in the generation of these two fields lies in the MSE that was chosen as the metric for the SHAP calculations. The first MSE reads

$$MSE_{\text{global}} = \frac{1}{4}\sum_{c=1}^{4} \mathbb{E}\left[(\hat{x}_{i,j,k,c} - x_{i,j,k,c})^2\right]_{i,j,k} \tag{21}$$

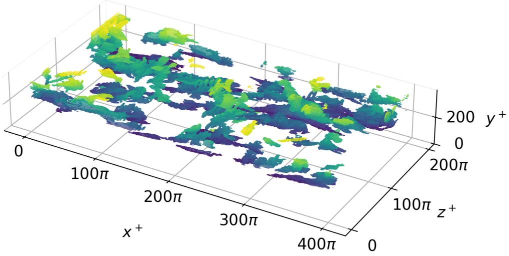
where the index $c = 1, 2, 3, 4$ indicates the channels, u , v, w, and T respectively, $x_{i,j,k,c}$ is the predicted output and $\hat{x}_{i,j,k,c}$ is the ground truth. The operator $\mathbb{E}\left[\,\cdot\,\right]_{i,j,k}$ represents the spatial averaging across the three spatial



(a)



(b)



(c)

Figure 9: Instantaneous visualization of the coherent structures in the channel flow. In the figure, the wall is located at $y^+ = 0$, and the mid-plane of the channel is $y^+ = 200$. The structures are colored depending on the distance from the wall, from purple (close to the wall) to yellow (close to the centerline). S1 structures are shown in subfigure (a), S2 structures are shown in subfigure (b), and S3 structures are shown in subfigure c). The structures shown are the ones in the lower half of the channel.

dimensions x, y, and z. Thus, the MSE is a scalar value representative of the overall accuracy of the predictions across all fields.
On the other hand, for the second SHAP field, the formulation of the MSE was changed to

$$MSE_T = \mathbb{E}\left[(\hat{x}_{i,j,k,4} - x_{i,j,k,4})^2\right]_{i,j,k} \tag{22}$$

only considering the overall error of the fourth channel, the temperature. This choice is motivated by physical causality. When the MSE includes both velocity and temperature errors, a clear asymmetry arises: temperature, treated as a passive scalar, does not influence the velocity field, whereas velocity directly governs temperature evolution. Consequently, the SHAP magnitudes associated with temperature are systematically smaller than those of the velocity components when the global MSE is used. This leads to identifying regions of importance where the temperature field has little to no influence, as will be shown in the statistics of the SHAP structures based on the first SHAP field. Furthermore, Q events and the thermal Q events are computed following the threshold criteria in Equation 17 and Equation 18, respectively.
We proceed by defining three types of SHAP structures were computed in order to evaluate the capability of the SHAP values to capture regions of interest. From this point forward, for the sake of clarity, we will refer to the different types of SHAP structures as S1, S2, and S3. S1 and S2 were computed from SHAP fields where the MSE, previously shown in Equation 21, accounted for the overall error of the velocity and temperature field. The first set of structures (S1) only took into account the velocity components of the SHAP field, and the criterion for the percolation reads

$$\sqrt{\phi_u^2(x,y,z,t) + \phi_v^2(x,y,z,t) + \phi_w^2(x,y,z,t)} > H_{S1}\sqrt{\overline{\phi_u^2}(y) + \overline{\phi_v^2}(y) + \overline{\phi_w^2}(y)}. \tag{23}$$

Conversely, the second set of structures (S2) was identified considering also ϕ_T, and its percolation condition is shown in Equation 19. The third set of structures (S3) was computed starting from the second type of SHAP field, where the MSE chosen was the one shown in Equation 22. We begin our analysis by comparing S1 ans S2, which allows the assessment of the impact of ϕ_T on structure segmentation in the case when a global MSE is considered. The JPDFs 10 of the velocity components conditioned on S1 and S2 are virtually indistinguishable across all y^+, indicating that inclusion of ϕ_T does not significantly modify the extracted regions. This can also be seen in a qualitative way in Figure 9, subfigures a) and b), where instantaneous representations of these

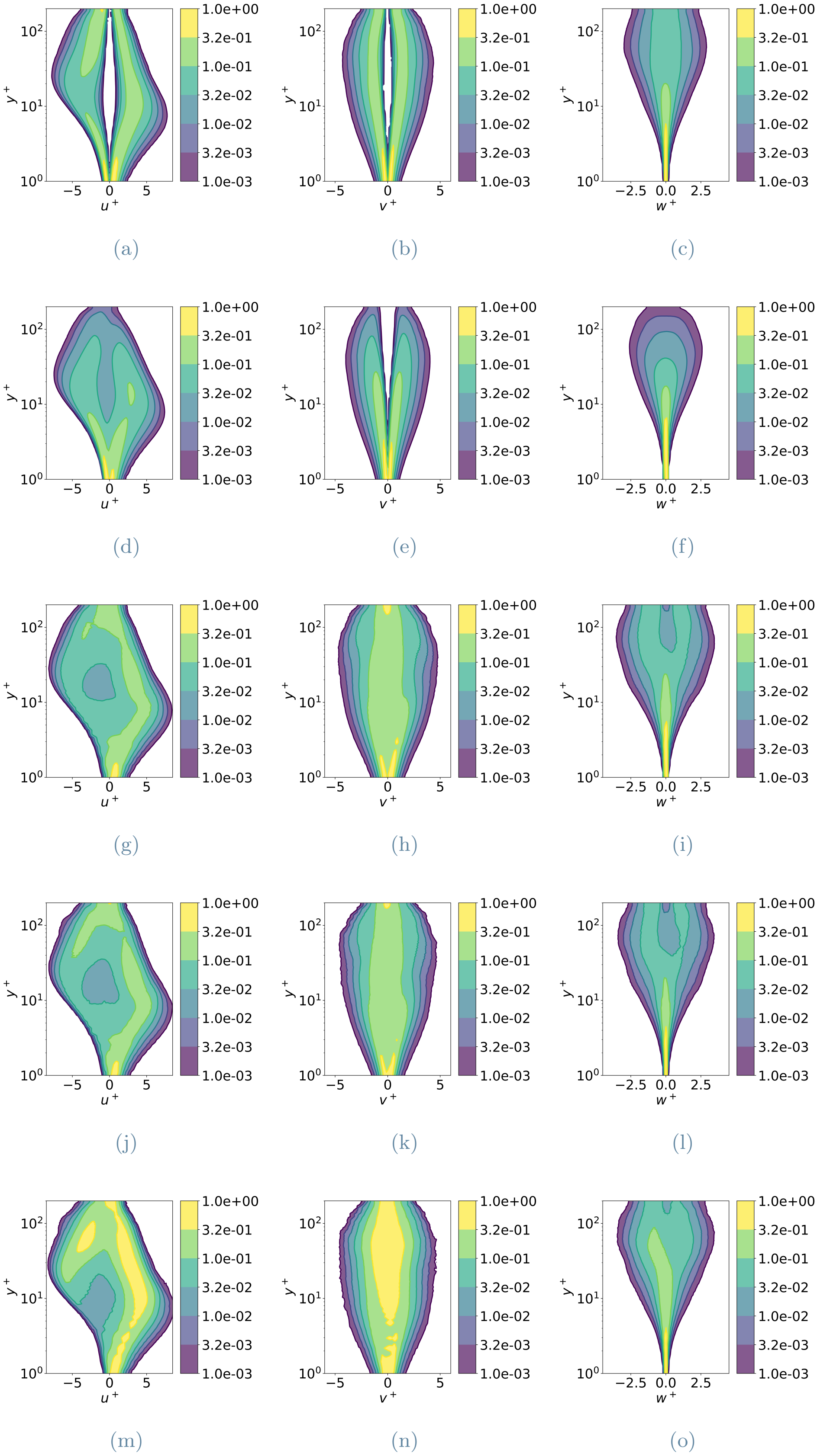


Figure 10: Joint probability density function of the three velocity fluctuation components, and the inner-scaled wall-normal distance, y^+, of the structures. The figure presents the five types of structures: the Reynolds-stress structures a), b), c), the thermal structures, subfigures d), e), and f), S1, g), h), i), S2, subfigures j), k), and l), and S3, subfigures m), n), and o).

two types of structures are shown. While the similarities of the JPDFs alone are insufficient to conclude that these SHAP structures are identical, further confirmation can be found in Figure 11, subfigure a), where the coincidence percentage in pairs between the three sets of structures S1, S2, and S3 is shown with respect to the wall-normal distance y^+. The pairwise coincidence exceeds 95% across the entire wall-normal range, confirming that ϕ_T has negligible influence when the global MSE is adopted. In the context of this data-driven framework, the SHAP component associated with temperature ultimately fails to meaningfully influence the extraction of influential regions when considered together with the other SHAP components. Consequently, when considering temperature as a passive scalar, other types of metrics need to be explored in order to gain insight from SHAP values with regard to heat transfer. However, even if ϕ_T was found to be uninfluential in the thresholding process for the case of S2, this does not imply that S1 and S2 cannot be informative in their own right: the presence of the temperature field error as a metric for the feature attribution does not only affect ϕ_T, but also the SHAP values associated with velocity components, since they are also predictive with respect to the temperature field evolution. This can be inferred by considering the previous validation case: When comparing the JPDF of the SHAP structures for the $Re_\tau = 125$ case (Figure 6), which did not include temperature in the loss definition, with the JPDF of S2 (Figure 10), noticeable differences emerge. These differences cannot be attributed to the change in friction Reynolds number. This is supported by the findings of [7], where SHAP structures obtained from channel flow simulations at $Re_\tau = 125$ and $Re_\tau = 550$ were directly compared, revealing only minor Reynolds-number effects on the resulting structures.

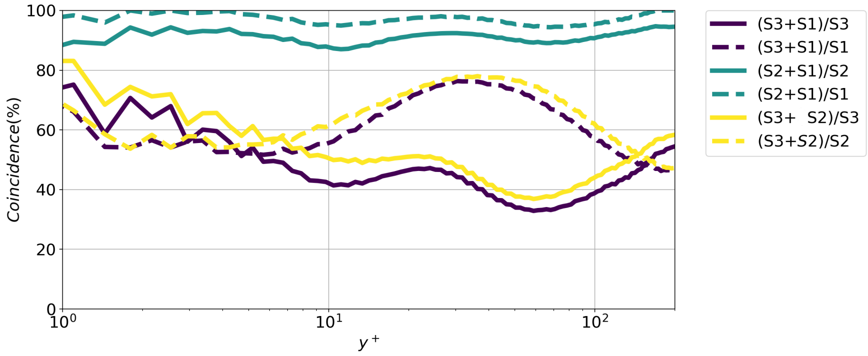
Therefore, the observed discrepancies are more plausibly linked to the inclusion of the temperature field error in the MSE. Even though temperature behaves as a passive scalar, incorporating its contribution into the MSE alters the importance attribution of ϕ_u, ϕ_v, and ϕ_w. In this sense, the velocity SHAP components become partially conditioned by their predictive relevance to heat transfer, leading to a modified structural selection.

Nonetheless, the global MSE introduces a structural weighting imbalance: three velocity components contribute to the loss, while temperature contributes only one. In a passive-scalar configuration where temperature does not influence momentum, this asymmetry guarantees that attribution will be dominated by velocity-driven dynamics. These results clearly indicate that the application of this framework must be tailored to the specific objective of the study. In particular, the definition of the MSE must reflect the underlying physical mechanisms governing the flow and the way these mechanisms influence the evolution of the state variables.

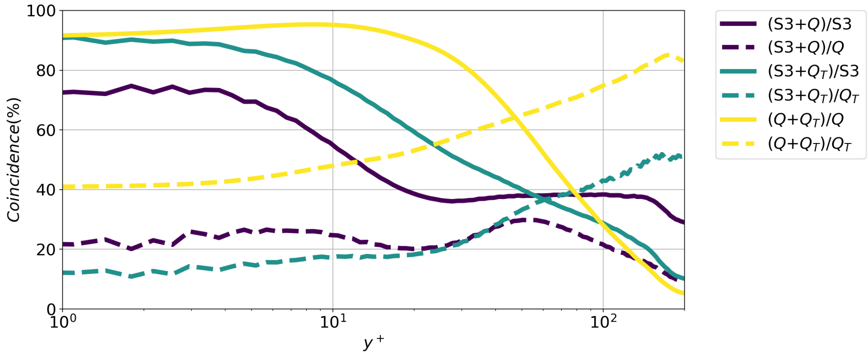
For this reason, a second SHAP field was computed using a loss function based solely on the temperature prediction error. Since the objective is to isolate the structures that most strongly influence temperature evolution, restricting the importance attribution to the thermal component allows the framework to target heat-transfer-relevant dynamics more directly.

Let us now examine the properties of the S3 structures. As shown in Figure 10, the JPDF of S3 displays a clear preference for positive streamwise fluctuations across the channel. Within the quadrant framework, these high-velocity states correspond predominantly to outward interactions (Q1: $v' > 0$) and sweeps (Q4: $v' < 0$). This indicates that S3 preferentially samples regions characterized by the presence of high-momentum fluid, particularly motions associated with the downward transport of such fluid toward the wall. In addition to this dominant feature, a secondary probability peak emerges at $y^+ \approx 80$ for $u^+ < 0$, a region commonly associated with ejections (Q2: $v' > 0$) and inward interactions (Q3: $v' < 0$). The location of this peak, within the logarithmic layer, suggests that S3 does not exclusively capture near-wall sweep-dominated dynamics, but also identifies structures linked to the upward transport of low-momentum fluid away from the wall. Compared with the classical Q event statistics, these velocity states appear significantly more concentrated within the S3 structures, highlighting the stronger statistical selectivity introduced by the SHAP. Rather than sampling all quadrant events uniformly, S3 isolates a subset of velocity configurations in which the magnitude and spatial coherence of the fluctuations are amplified. This concentration indicates that the S3 structures preferentially identify locations where coupled momentum–heat exchange mechanisms are most dynamically relevant. The predominance of positive streamwise fluctuations suggests that S3 is preferentially linked to the downward transport of high-momentum fluid toward the wall. However, the emergence of the secondary peak in the logarithmic region indicates that the dynamical role of S3 evolves with wall-normal distance. Rather than representing a single coherent motion, the dual-peak topology, consisting of a primary ridge at $u^+ > 0$ and a secondary island at $u^+ < 0$, suggests that S3 captures an interfacial region where high-speed sweeps and low-speed ejections coexist and interact.

While the JPDFs are helpful in exploring the data enclosed in the identified structures, they are not enough to fully describe them. Hence, to further characterize the structures from a spatial distribution perspective, the coincidence of the structures normalized by the respective occupied volumes as a function of the inner-scaled wall-normal distance y^+ was computed, as illustrated in Figure 11. In subfigure a), the coincidences (evaluated in pairs) of S1, S2, and S3 are depicted. As established earlier, the coincidence of S1 and S2 is almost perfect across all values of y^+. With respect to S3 and S2, the coincidence is quite high, ranging from 40% at the lowest to 80% at the highest when considering the normalization done through the volume of S3. This further reinforces the conclusion that, at least in part, information regarding the temperature field and the heat transfer is already embedded in the SHAP components associated with the velocity components, which is consistent with



(a)



(b)

Figure 11: Coincidence between the various types of structures. S1, S2, and S3, subfigure a). S3, Reynolds-stress structures (Q), and thermal structure (Q_T)

the underlying physics of the passive scalar advection.
In subfigure b), the coincidences (evaluated in pairs) between S3, Q events, and the thermal Q events is portrayed. Firstly, it is possible to see from the volume normalization that, unlike the previous case where the SHAP structures were compared, the occupied volumes of the structures are very different: in particular, in the interval of $y^+ < 50$ the thermal structures occupy the most space compared to the rest, while for $y^+ > 50$ the Q events are the ones that take up the most volume in the channel. Secondly, the coincidences of S3 with respect to the rest of the structures suggest that the regions identified by the SHAP structures are also occupied by the classical structures, particularly in the $y^+ < 10$ range. For $y^+ > 10$, the coincidence $S3 + Q/S3$ descends and stabilizes at around 40%, and $S3 + Q_T/S3$ keeps steadily descending. This indicates that the SHAP methodology allocates importance to different regions of the flow compared to classical structures, such as the flow near the centerline, where turbulent kinetic energy production and dissipation are not as pronounced as they are in the near-wall regions. While classical coherent structures remain effective for characterizing near-wall turbulence dynamics, the SHAP-based methodology highlights additional regions of statistical relevance, particularly within the bulk of the flow. In passive-scalar configurations, a temperature-specific MSE is required to isolate heat-transfer-relevant structures. These results demonstrate that interpretability in data-driven models is not invariant to the choice of the metric used for the SHAP computation. Consequently, the physical insights derived from such frameworks are conditional on selecting an objective function that aligns with the underlying physics of interest.

4. Conclusions

This work presented an comprehensive investigation into the use of explainable deep learning for identifying dynamically important regions in turbulent channel flows, with particular emphasis on the interaction between velocity dynamics and passive scalar transport. By combining direct numerical simulation data with a three-dimensional U-Net surrogate model and an explainable attribution framework based on Gradient-SHAP, the study established a pipeline for mapping predictive relevance onto physical flow structures. The resulting methodology moves beyond purely kinematic descriptions of turbulence and introduces a data-driven perspective centered on short-time dynamical influence.

The first major outcome of the study is the successful validation of the deep-learning framework on a canonical smooth channel flow. The neural network achieved low normalized prediction errors across all velocity components, demonstrating its capability to learn short-time nonlinear flow evolution with sufficient fidelity to support attribution analysis. The SHAP-derived importance fields reproduced the main statistical features reported in prior studies and showed meaningful correspondence with classical Reynolds-stress structures. This agreement confirms that the explainability framework is not merely highlighting numerical artifacts of the surrogate model, but is instead capturing physically interpretable regions associated with established turbulence mechanisms.

The extension of the framework to a turbulent channel flow with temperature treated as a passive scalar constitutes the central contribution of this work. In this configuration, the analysis exposed an inherent asymmetry in the coupling between velocity and scalar fields: while the velocity field strongly influences scalar evolution, the passive scalar does not feed back into the momentum equations. When SHAP values were computed using a combined multi-channel loss function, the resulting structures were dominated by velocity-driven importance, closely resembling those obtained in the purely hydrodynamic case. This result highlights a critical methodological insight: attribution methods in multi-physics systems are sensitive to the definition of the objective function, and naive aggregation of prediction errors can obscure physically meaningful contributions from weaker but still relevant variables.

To address this limitation, alternative SHAP formulations focused specifically on scalar prediction were introduced. These temperature-centered structures highlight additional dynamically relevant regions beyond those emphasized by traditional diagnostics from classical momentum-based coherent structures. Although partial overlap with traditional Reynolds-stress and scalar-flux structures was observed, the SHAP-based structures also identified regions that are not captured by standard threshold-based criteria. This finding suggests that explainable deep learning can uncover influential flow regions that are less emphasized by conventional instantaneous criteria, particularly in systems involving coupled transport processes.

A key conceptual implication of the results is that predictive importance provides a complementary lens through which turbulence can be interpreted. Classical coherent-structure analysis emphasizes instantaneous transport quantities such as Reynolds stresses and scalar fluxes, whereas the SHAP framework ranks regions according to their contribution to short-time dynamical evolution as learned by the surrogate model. The partial mismatch between these perspectives indicates that dynamically influential regions are not always identical to those highlighted by traditional kinematic measures. This distinction is especially relevant for applications in flow control and model reduction, where the objective is to identify regions whose manipulation would most strongly affect future system behavior.

From a methodological standpoint, the study also provides practical guidelines for applying explainable attribution techniques to high-dimensional fluid dynamics problems. The percolation-based structure identification procedure proved effective in isolating discrete importance regions from noisy attribution fields, and the sensitivity analysis used to select threshold parameters ensured a balanced representation of coherent events. Furthermore, the use of periodic shifts to smooth SHAP fields leveraged the physical symmetries of the flow to reduce stochastic noise without introducing artificial bias. Together, these elements form a reproducible workflow for extracting interpretable structures from neural-network predictions of turbulent flows.

Despite these advances, several limitations and avenues for future research remain. The present analysis is restricted to moderate Reynolds numbers and passive scalar transport in a canonical channel geometry. Extending the framework to higher Reynolds numbers would test its scalability and its ability to capture increasingly complex multiscale interactions. Incorporating active scalar effects, such as buoyancy-driven coupling, would further challenge the attribution methodology by introducing bidirectional physical feedback. In addition, applying the framework to more complex geometries and boundary conditions could assess its robustness in practical engineering configurations. On the machine-learning side, exploring alternative architectures and attribution methods may improve both predictive accuracy and interpretability.

In summary, this work demonstrates that explainable deep learning can serve as a powerful tool for identifying and characterizing dynamically relevant structures in turbulent flows with passive scalar transport. By bridging data-driven modeling and physical interpretation, the proposed framework enriches the analysis of turbulence beyond traditional coherent-structure paradigms. The results indicate that attribution-based structures capture aspects of flow dynamics that complement classical diagnostics and may inform future strategies for turbulence modeling, control, and optimization. As data-driven methods continue to integrate with computational fluid dynamics, such interpretable frameworks will play an increasingly important role in translating predictive performance into physical understanding.

References

[1] G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25:539–575, 1993.

[2] Steven L Brunton, Bernd R Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual review of fluid mechanics*, 52(1):477–508, 2020.

[3] D. R. Carlson and J. L. Lumley. Quadrant analysis of turbulent heat flux in a channel flow. *International Journal of Heat and Mass Transfer*, 37:2593–2604, 1994.

[4] Jurgita Černevičienė and Audrius Kabašinskas. Explainable artificial intelligence (xai) in finance: a systematic literature review. *Artificial Intelligence Review*, 57(8):216, 2024.

[5] Pedro Costa, Everett Phillips, Luca Brandt, and Massimiliano Fatica. Gpu acceleration of cans for massively-parallel direct numerical simulations of canonical fluid flows. *Computers & Mathematics with Applications*, 81:502–511, 2021.

[6] Andrés Cremades, Sergio Hoyas, Rahul Deshpande, Pedro Quintero, Martin Lellep, Will Junghoon Lee, Jason P. Monty, Nicholas Hutchins, Moritz Linkmann, Ivan Marusic, and Ricardo Vinuesa. Identifying regions of importance in wall-bounded turbulence through explainable deep learning. *Nature Communications*, 15(1), May 2024.

[7] Andrés Cremades, Sergio Hoyas, and Ricardo Vinuesa. Classically studied coherent structures only paint a partial picture of wall-bounded turbulence, 2025.

[8] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016.

[9] Gabriel Erion, Joseph D Janizek, Pascal Sturmfels, Scott M Lundberg, and Su-In Lee. Improving performance of deep learning models with axiomatic attribution priors and expected gradients. *Nature machine intelligence*, 3(7):620–631, 2021.

[10] Uriel Frisch. *Turbulence: The Legacy of A. N. Kolmogorov*. Cambridge University Press, 1995.

[11] Kai Fukami, Koji Fukagata, and Kunihiko Taira. Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics*, 870:106–120, 2019.

[12] Luca Guastoni, Miguel P. Encinar, Philipp Schlatter, Sergio Hoyas, and Ricardo Vinuesa. Convolutional-network models to predict wall-bounded turbulence from wall measurements. *Journal of Fluid Mechanics*, 928:A27, 2021.

[13] Geoffrey Hinton. Neural networks for machine learning, lecture 6e. `https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`, 2012. Coursera lecture slides.

[14] J. C. R. Hunt, A. A. Wray, and P. Moin. Eddies, streams, and convergence zones in turbulent flows. *Center for Turbulence Research Report CTR-S88*, 1988.

[15] J. Jeong and F. Hussain. On the identification of a vortex. *Journal of Fluid Mechanics*, 285:69–94, 1995.

[16] Nobuhide Kasagi, Yoji Tomita, and A. Kuroda. Direct numerical simulation of passive scalar transport in a turbulent channel flow. *Journal of Heat Transfer*, 114(3):598–606, 1992.

[17] H. T. Kim, S. J. Kline, and W. C. Reynolds. The production of turbulence near a smooth wall in a turbulent boundary layer. *Journal of Fluid Mechanics*, 50:133–160, 1971.

[18] John Kim, Parviz Moin, and Robert Moser. Turbulence statistics in fully developed channel flow at low reynolds number. *Journal of Fluid Mechanics*, 177:133–166, 1987.

[19] P. S. Klebanoff. Turbulence characteristics of a boundary layer and their relation to the viscous sublayer. *NACA Report*, 1247, 1962.

[20] S. J. Kline, W. C. Reynolds, F. A. Schraub, and P. W. Runstadler. The structure of turbulent boundary layers. *Journal of Fluid Mechanics*, 30:741–773, 1967.

[21] A. N. Kolmogorov. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high reynolds number. *Journal of Fluid Mechanics*, 13(1), 1962.

[22] Andrei Nikolaevich Kolmogorov, V. Levin, Julian Charles Roland Hunt, Owen Martin Phillips, and David Williams. The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 434(1890):9–13, 1991.

[23] T. von Kármán. The fundamentals of the statistical theory of turbulence. *Journal of the Aeronautical Sciences*, 4(4):131–138, 1937.

[24] Adrián Lozano-Durán, Oscar Flores, and Javier Jiménez. The three-dimensional structure of momentum transfer in turbulent channels. *Journal of Fluid Mechanics*, 694:100–130, 2012.

[25] S. S. Lu and W. W. Willmarth. Measurements of the structure of the reynolds stress in a turbulent boundary layer. *Journal of Fluid Mechanics*, 60:481–511, 1973.

[26] J. L. Lumley. The structure of inhomogeneous turbulent flows. *Atmospheric Turbulence and Radio Wave Propagation*, pages 166–178, 1967.

[27] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[28] Dang Minh, H Xiang Wang, Y Fen Li, and Tan N Nguyen. Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, 55(5):3503–3568, 2022.

[29] Ryuichi Nagaosa and Robert A Handler. Statistical analysis of coherent vortices near a free surface in a fully developed turbulence. *Physics of Fluids*, 15(2):375–394, 2003.

[30] M. Nagata. Three-dimensional finite-amplitude solutions in plane couette flow: bifurcation from infinity. *Journal of Fluid Mechanics*, 217:519–527, 1990.

[31] Roddam Narasimha, S Rudra Kumar, A Prabhu, and SV Kailas. Turbulent flux events in a nearly neutral atmospheric boundary layer. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1852):841–858, 2007.

[32] Lewis Fry Richardson. *THE FUNDAMENTAL EQUATIONS*, page 21–114. Cambridge Mathematical Library. Cambridge University Press, 2007.

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[34] A. Segalini and P. H. Alfredsson. Quadrant analysis of heat and momentum flux in turbulent channel flow. *Physics of Fluids*, 26:025109, 2014.

[35] S. C. Shadden, F. Lekien, and J. E. Marsden. Definition and properties of lagrangian coherent structures from finite-time lyapunov exponents. *Physica D*, 212:271–304, 2005.

[36] Lloyd S Shapley et al. A value for n-person games. 1953.

[37] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR, 2017.

[38] G. I. Taylor. The spectrum of turbulence. *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences*, 164(919):476–490, 1938.

[39] Geoffrey Ingram Taylor. Statistical theory of turbulenc. *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences*, 151(873):421–444, 1935.

[40] Nils Thuerey, Konstantin Weißenow, Lukas Prantl, and Xiangyu Hu. Deep learning methods for reynolds-averaged navier–stokes simulations of airfoil flows. *AIAA journal*, 58(1):25–36, 2020.

[41] Ricardo Vinuesa and Steven L Brunton. Enhancing computational fluid dynamics with machine learning. *Nature Computational Science*, 2(6):358–366, 2022.

[42] F. Waleffe. On a self-sustaining process in shear flows. *Physics of Fluids*, 9:883–900, 1997.

[43] 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.

[44] Zellman Warhaft. Passive scalars in turbulent flows. *Annual Review of Fluid Mechanics*, 32(1):203–240, 2000.

[45] WW Willmarth and SS Lu. Structure of the reynolds stress near the wall. *Journal of Fluid Mechanics*, 55(1):65–92, 1972.

Abstract in lingua italiana

Il presente lavoro tratta di identificazione di regioni dinamicamente rilevanti in flussi turbolenti in canale piano mediante l'impiego di tecniche di explainable deep learning, con un'attenzione specifica all'interazione tra il trasporto della quantità di moto e la dinamica di uno scalare passivo. Una rete neurale convoluzionale U-Net tridimensionale è stata addestrata su dati di simulazione numerica diretta (DNS) per predire l'evoluzione temporale a breve termine dei campi di moto. La spiegabilità del modello è introdotta attraverso il framework Gradient-SHAP, che associa a ogni punto della griglia computazionale un valore di importanza basato sul contributo all'errore di predizione. Questi campi di importanza permettono di definire le cosiddette strutture basate su SHAP tramite una procedura di percolazione, consentendo una caratterizzazione oggettiva e data-driven delle regioni del flusso che esercitano la maggiore influenza sull'evoluzione temporale.
La metodologia è stata preliminarmente validata su un caso canonico di canale turbolento, riproducendo risultati noti in letteratura e garantendo errori di predizione contenuti e statistiche strutturali consistenti. Successivamente, l'analisi è stata estesa a un flusso turbolento in cui la temperatura è trattata come uno scalare passivo. Sono state esaminate diverse definizioni di strutture SHAP per valutare come differenti formulazioni della loss function influenzino l'identificazione delle zone dinamicamente significative. Il confronto tra le strutture SHAP e le strutture coerenti classiche (identificate tramite criteri basati sugli sforzi di Reynolds e sui flussi turbolenti di calore) è stato condotto mediante analisi di densità di probabilità congiunta (JPDF).
I risultati indicano che le strutture basate su SHAP si sovrappongono parzialmente a quelle tradizionali, rivelando tuttavia ulteriori regioni di interesse non identificate da criteri puramente cinematici. Nel caso dello scalare passivo, l'analisi evidenzia un'asimmetria nell'influenza causale tra i campi di velocità e di temperatura, dimostrando come formulazioni della SHAP focalizzate specificamente sulla temperatura siano indispensabili per isolare le strutture termicamente rilevanti. In conclusione, lo studio stabilisce un framework interpretabile e guidato dai dati per l'identificazione di strutture turbolente influenti, fornendo linee guida metodologiche per la sua applicazione a problemi di trasporto accoppiato.

Parole chiave: Turbolenza, Machine learning, Valori Shapley, Spiegabilità, Strutture coerenti