

Unsupervised identification of motifs in turbulent channel flow using Latent Dirichlet Allocation

Mohamed Frihat, Nemo Malhomme, Bérengère Podvin, Lionel Mathelin, Yann Fraigneau
and François Yvon

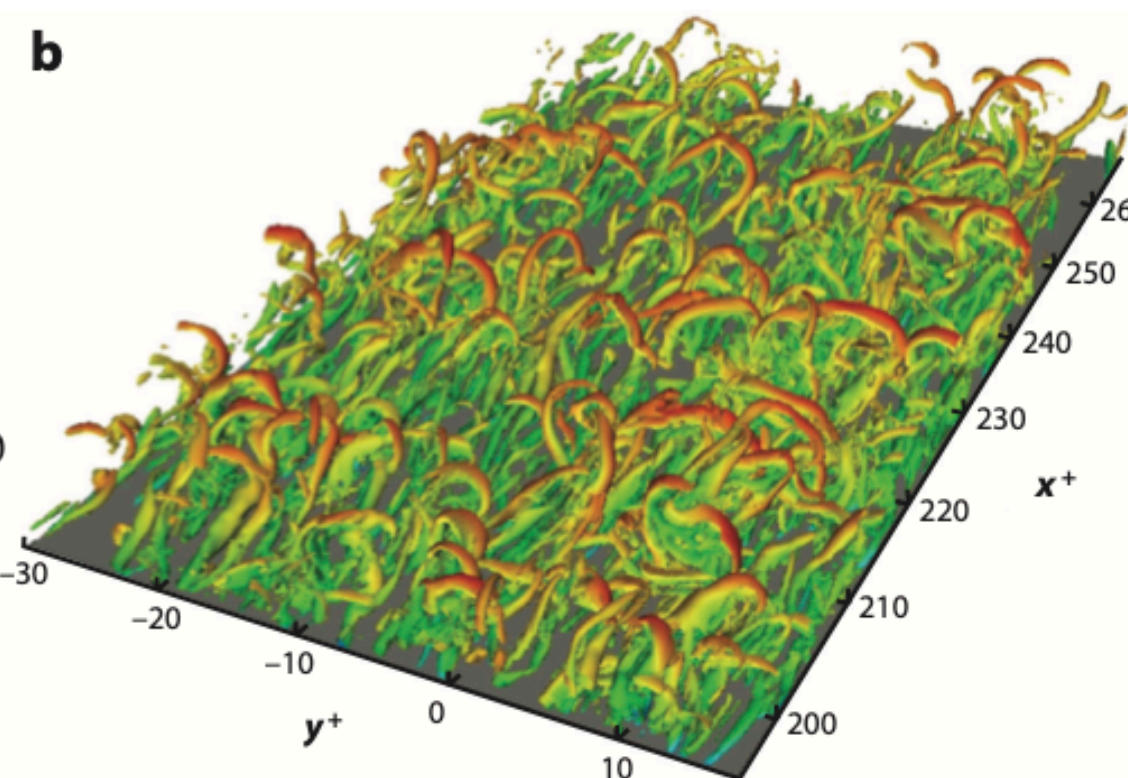
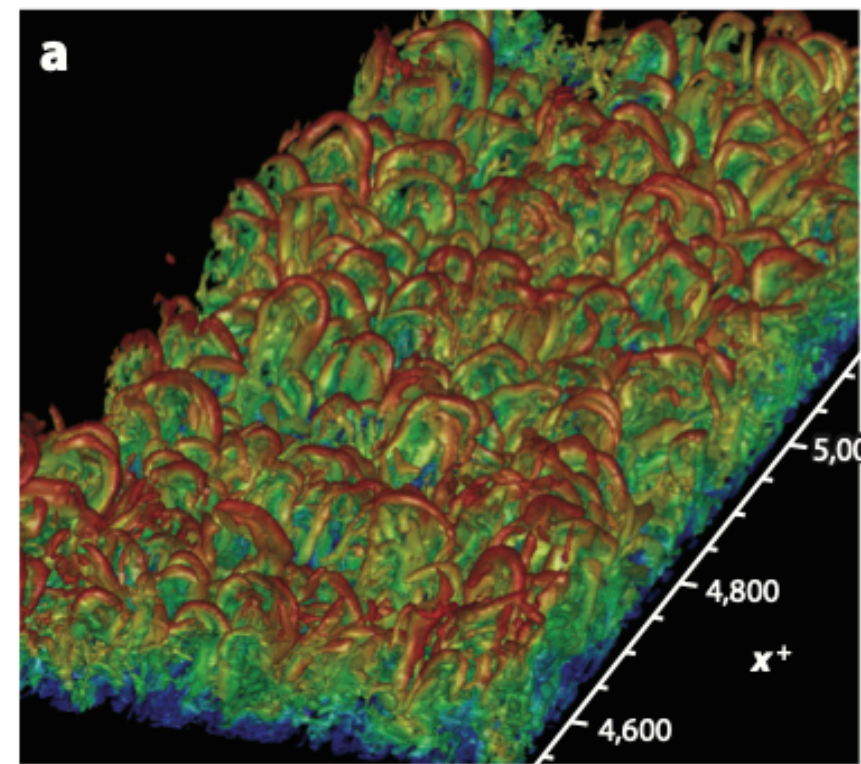
LISN, CNRS, Université Paris-Saclay

Turbulence theory

The search for coherent structures

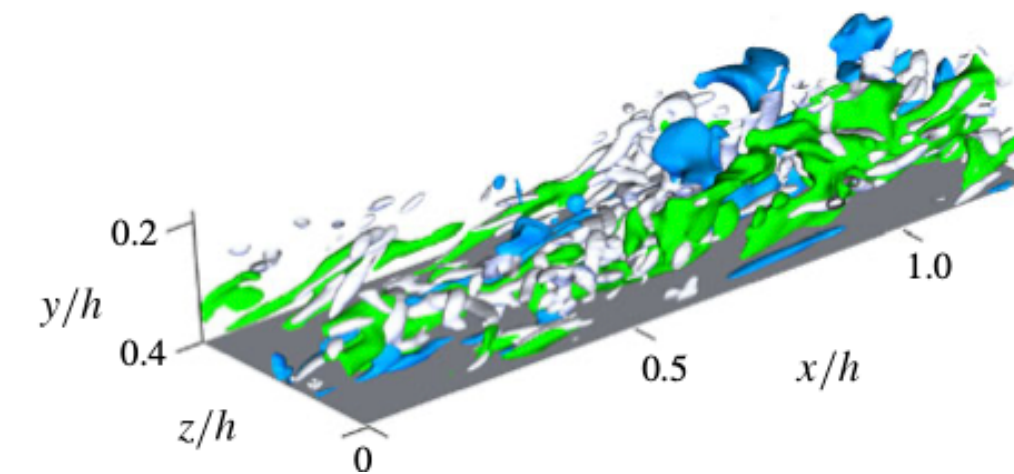
Hairpin or no hairpin?

(Monty and Marusic 2019)

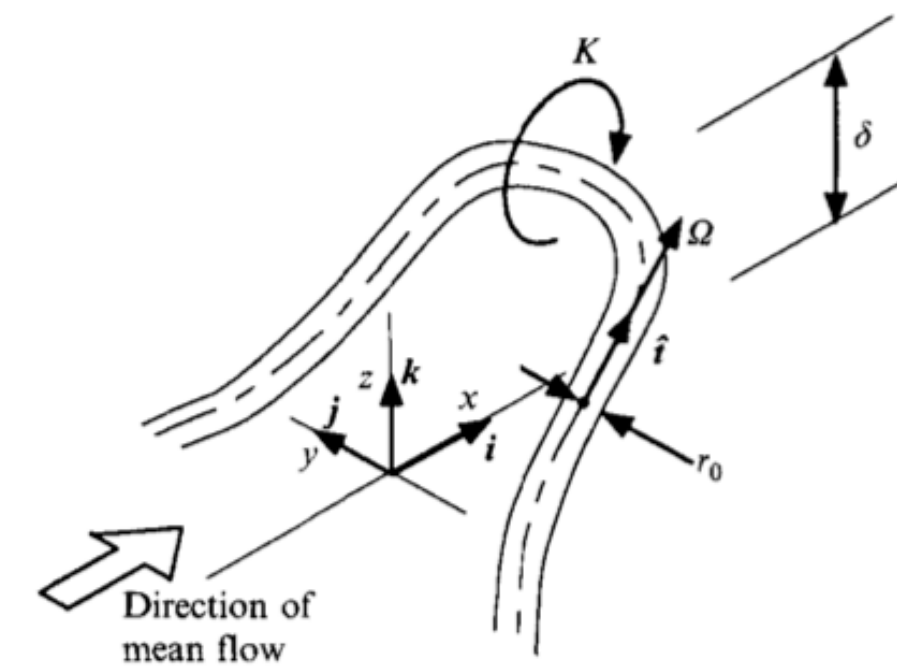


Wang and Moin 2009

Eitel et al. 2015



Q- events (Lozano-Duran et al. 2012)



Eddy-attached model (Perry and Marusic 1995)

Motivation

The search for coherent structures with machine learning

- Increase in data from simulations and experiments
- Need for techniques that are able to extract relevant information from data
- Existing unsupervised classification methods: Proper Orthogonal Decomposition (POD), Dynamic Mode Decomposition (DMD), k-means...
- ...Allow identification of latent factors but lack an inherent probabilistic framework

Looking for latent factors

A familiar concept in natural language processing

- Classification to extract **meaning** from a collection of documents containing words

	Collection	Observations	Key	Latent Factors
NLP	Documents	Words	M_{jk} = Number of times word j appears in document k	Topics n
Fluid Mechanics	Snapshots	Cells	M_{jk} = Intensity measured at cell j in snapshot k	Motifs n

How does LDA relate to other techniques?

★ POD = LSA (Latent Semantic Allocation)

- LSA \rightarrow pLSA = LSA with a motif-cell distribution
- pLSA \rightarrow LDA = pLSA with a snapshot-motif distribution

★ k-means \leftrightarrow hard clustering: each snapshot is assigned to one cluster

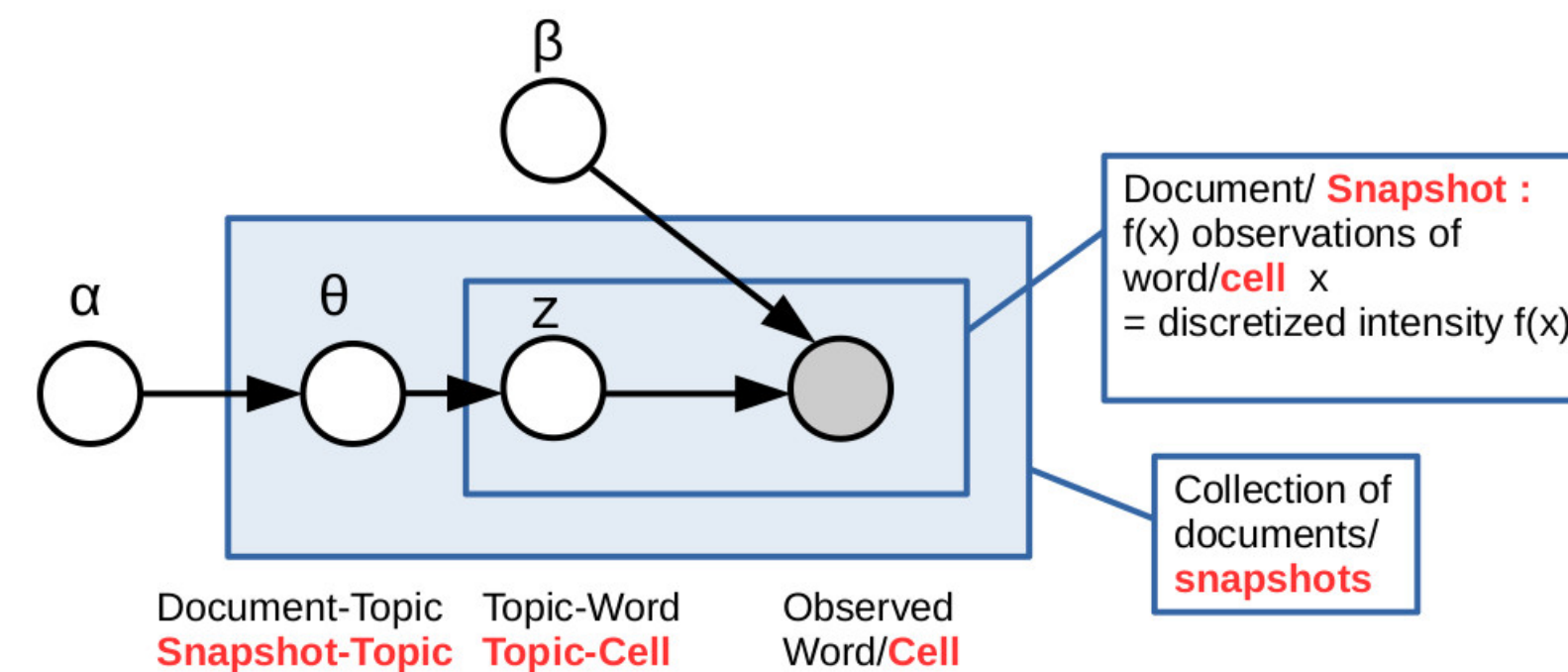
- LDA **soft** clustering technique with a generative model

What is a motif?

Latent Dirichlet Allocation (Blei et al. 2003)

Motifs are defined through two kinds of distributions:

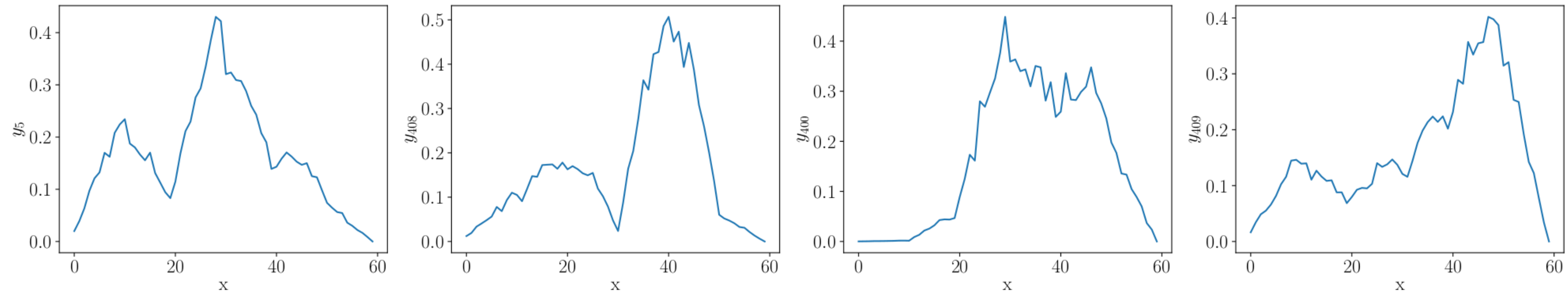
- How each motif n is distributed on the cells j (*spatial mode*) $\varphi_{n,j}$
- How each snapshot k is distributed on the motifs n (*amplitude*) $\theta_{k,n}$



LDA: φ and θ are modelled with **Dirichlet** distributions of parameter α and β

An example

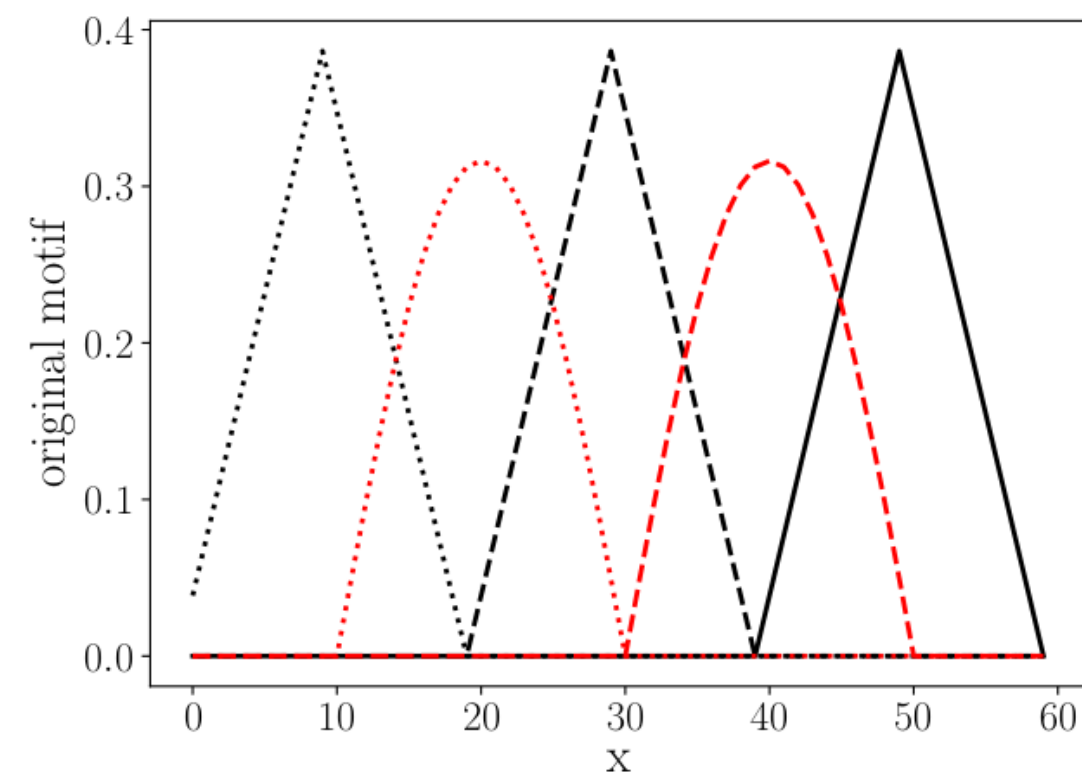
Snapshots



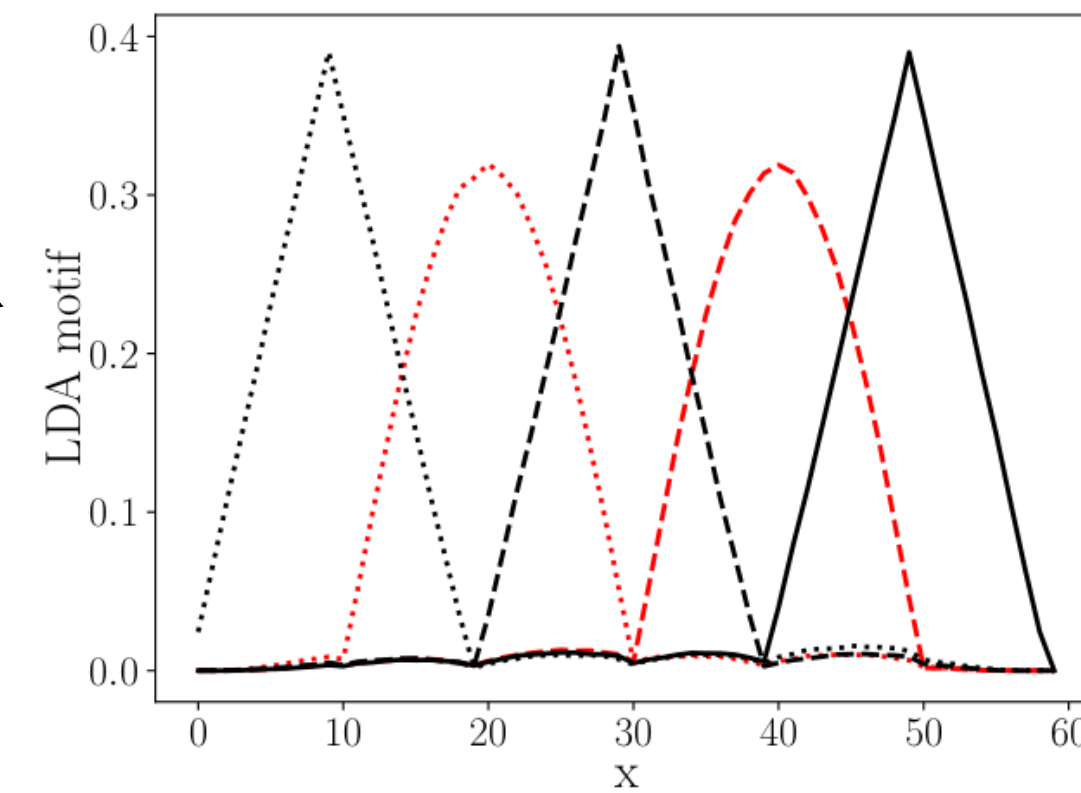
Creation 

 LDA

Original Motifs

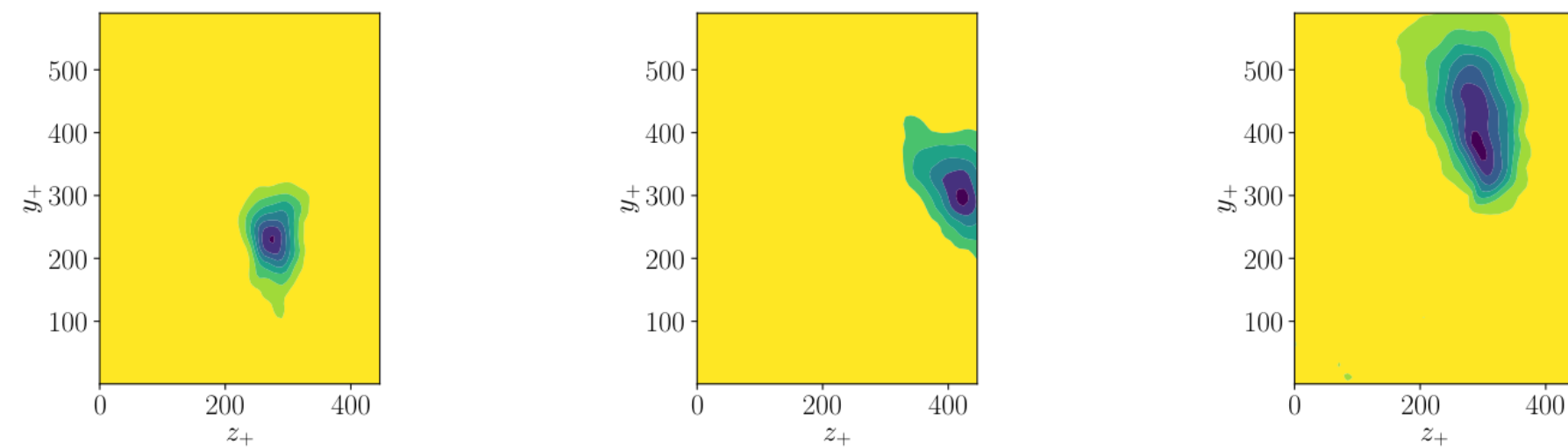
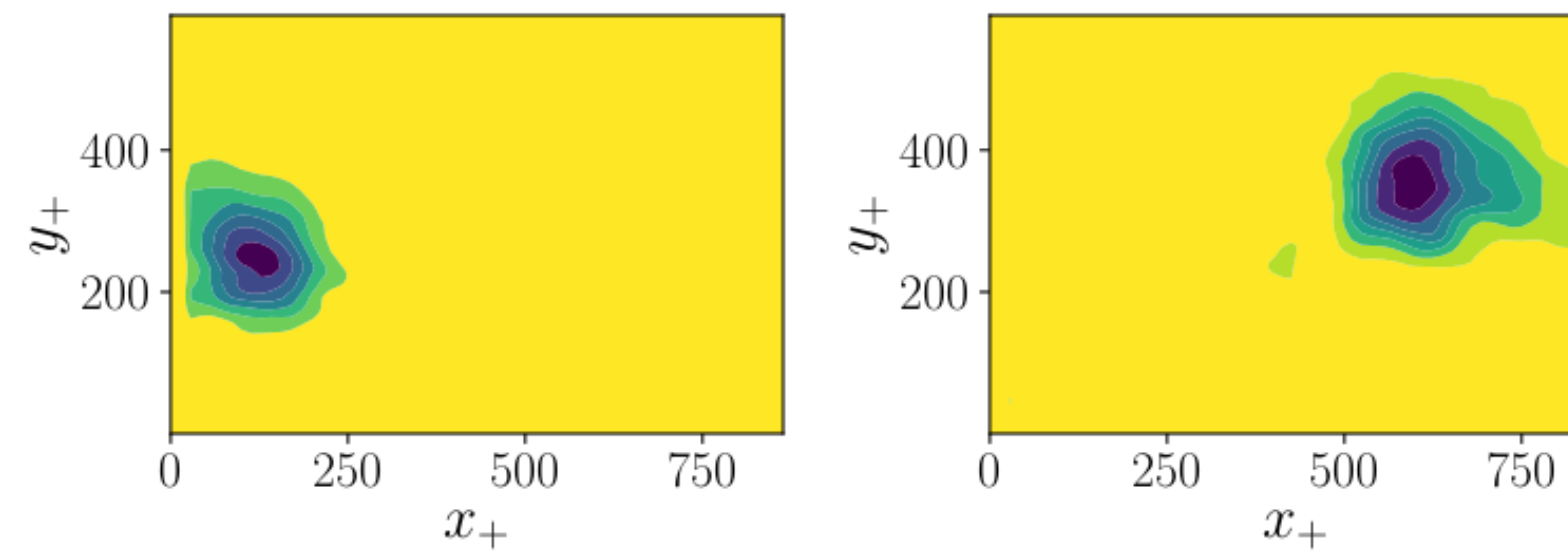
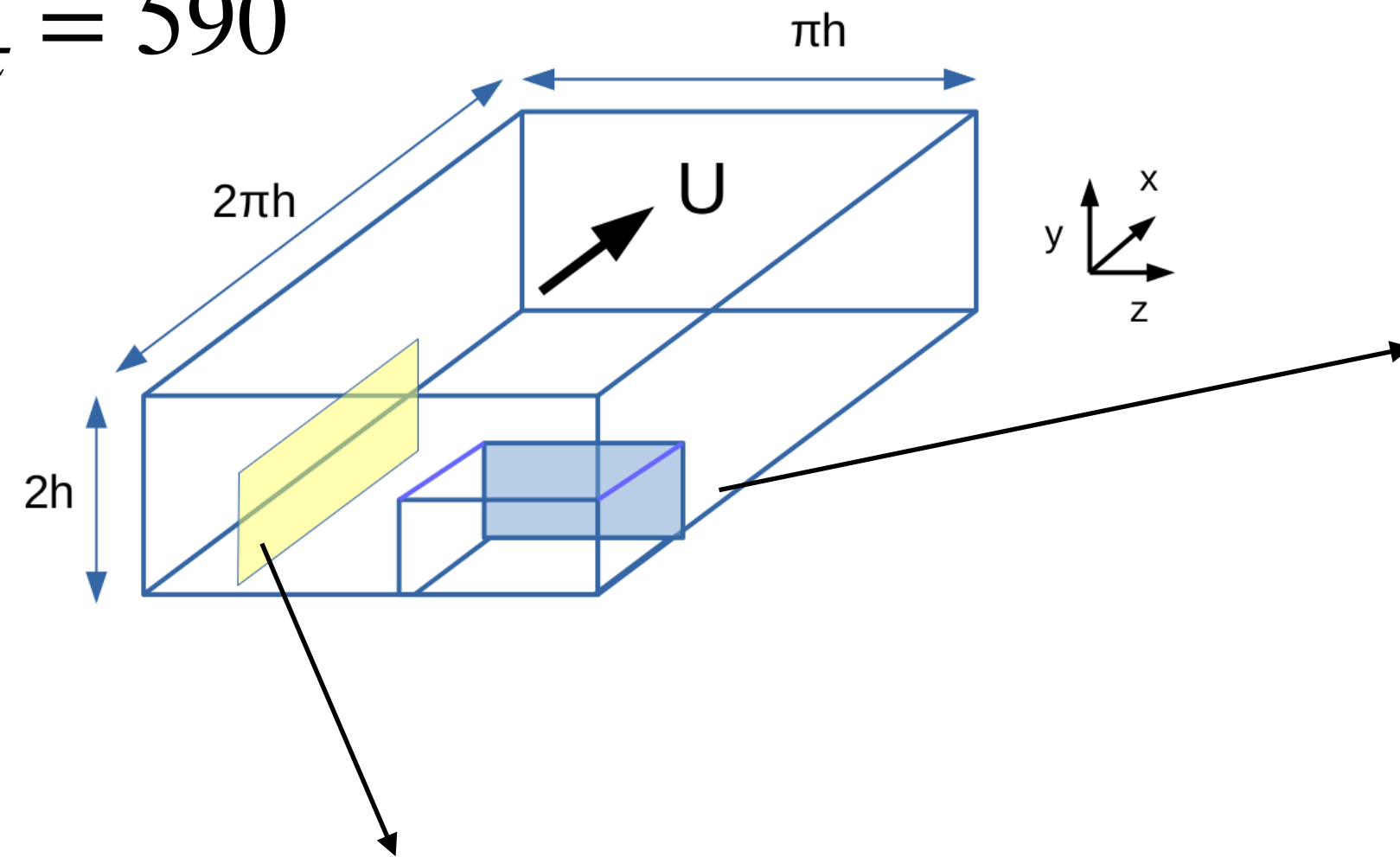


Recovered Motifs



Application to turbulent channel flow

$$R_\tau = 590$$

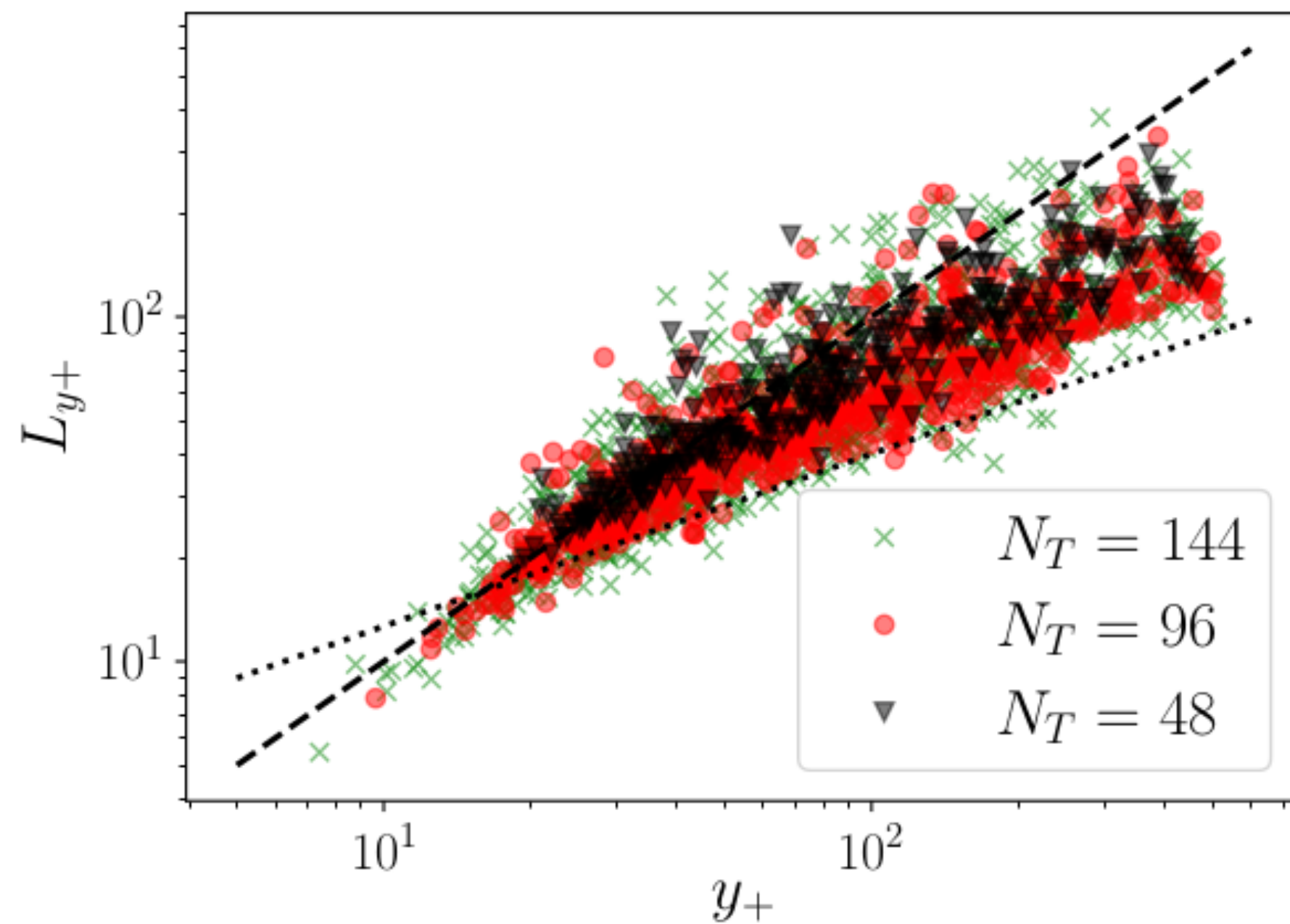


Motif-cell distribution
of instantaneous Q- Reynolds stress

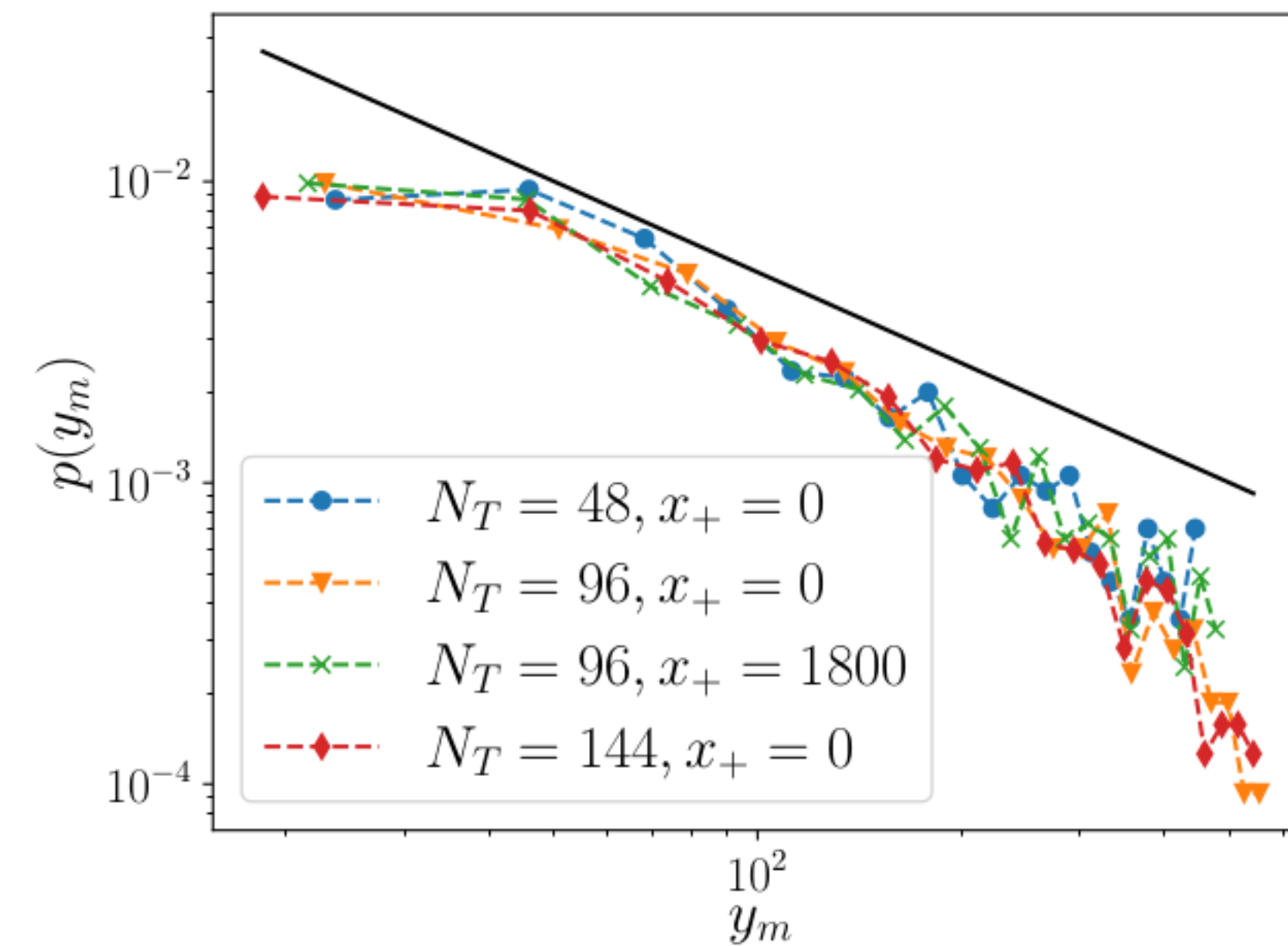
Do motifs scale with wall distance?

Confronting motifs with eddy attached (or detached!) hypotheses

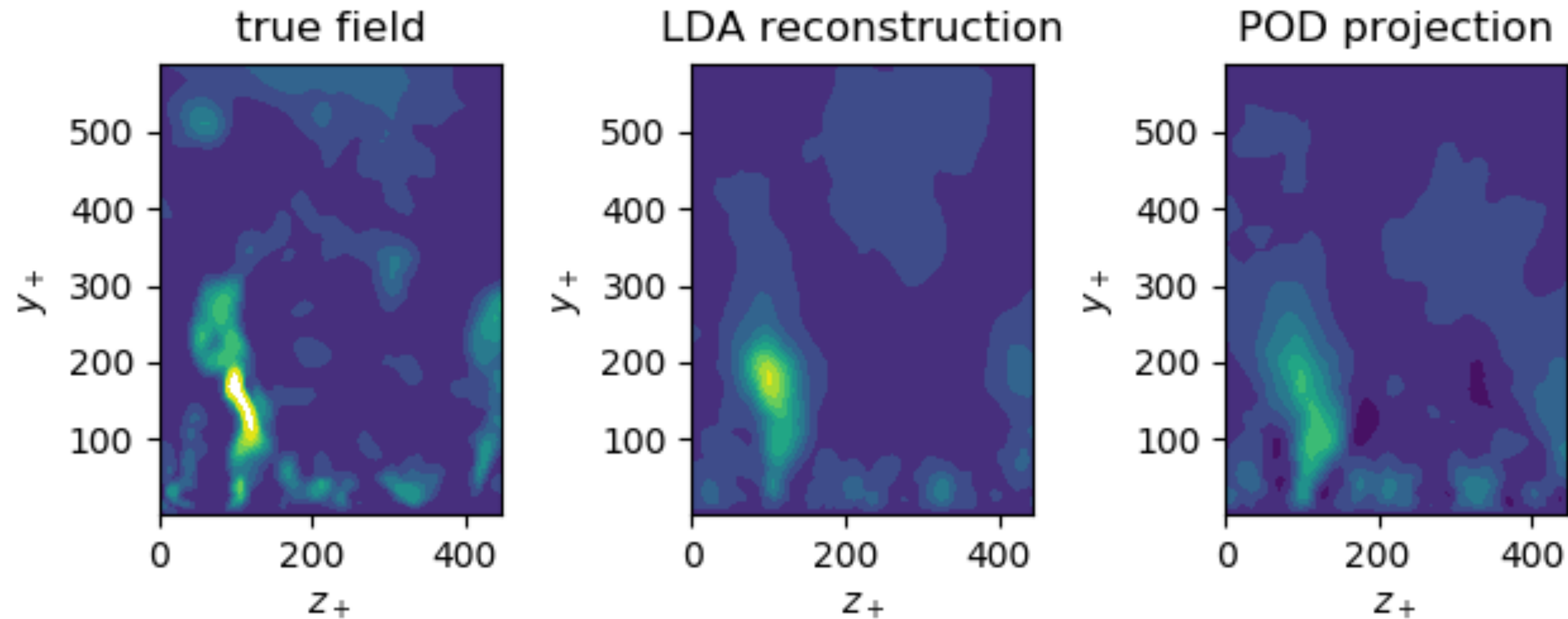
Linear scaling of motif size



Inverse scaling of numbers of motifs



Using LDA to reconstruct snapshots



Snapshot $\tau_{-,jk}$

R-LDA

R-POD

Cell x_j

Topics φ_n

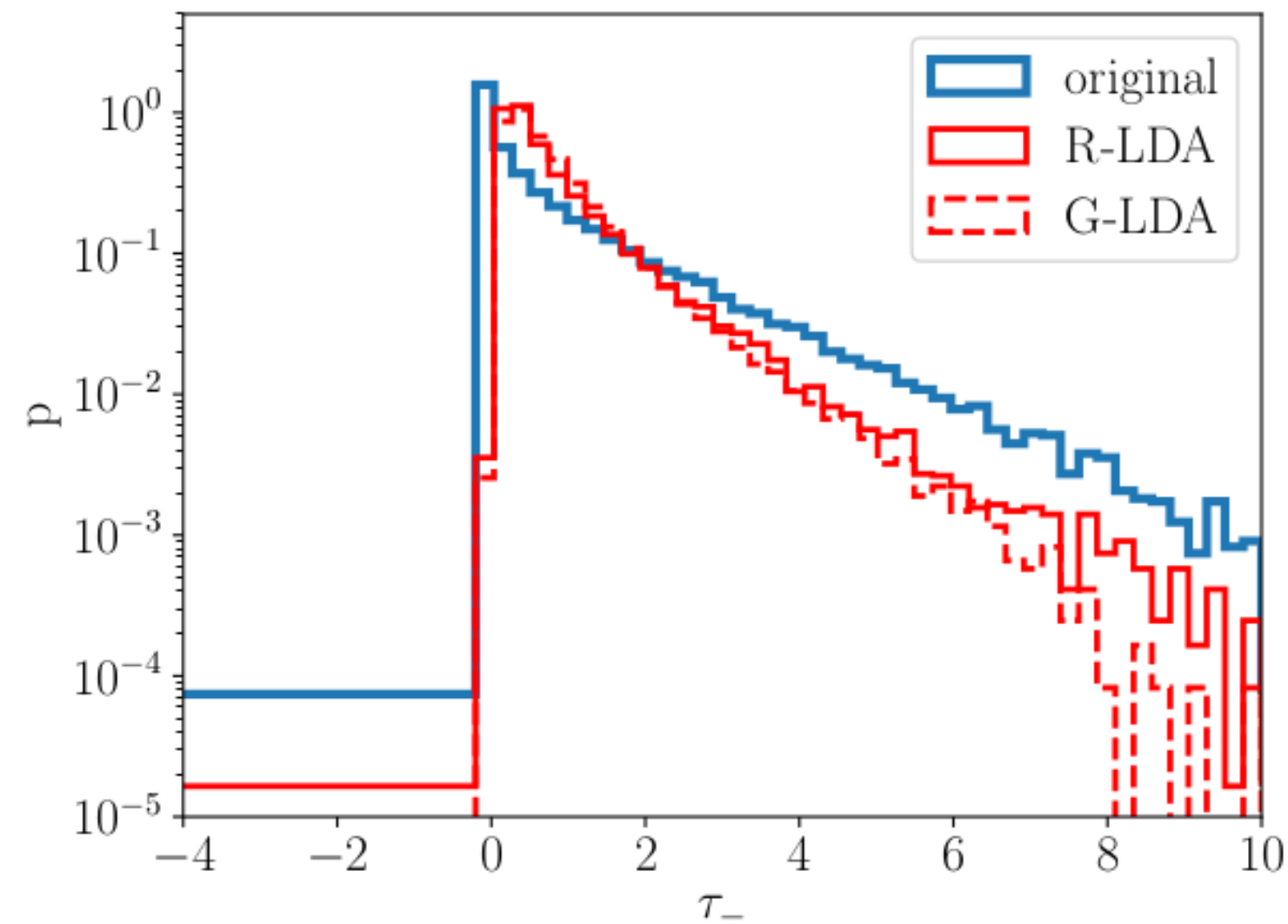
Modes ϕ_n

$$\tau_{-,jk}^{LDA} = \sum_n \theta_{kn} \varphi_{nj}$$

$$\tau_{-,jk}^{POD} = \sum_n a_{nk} \phi_n(x_j)$$

Using LDA to generate new snapshots

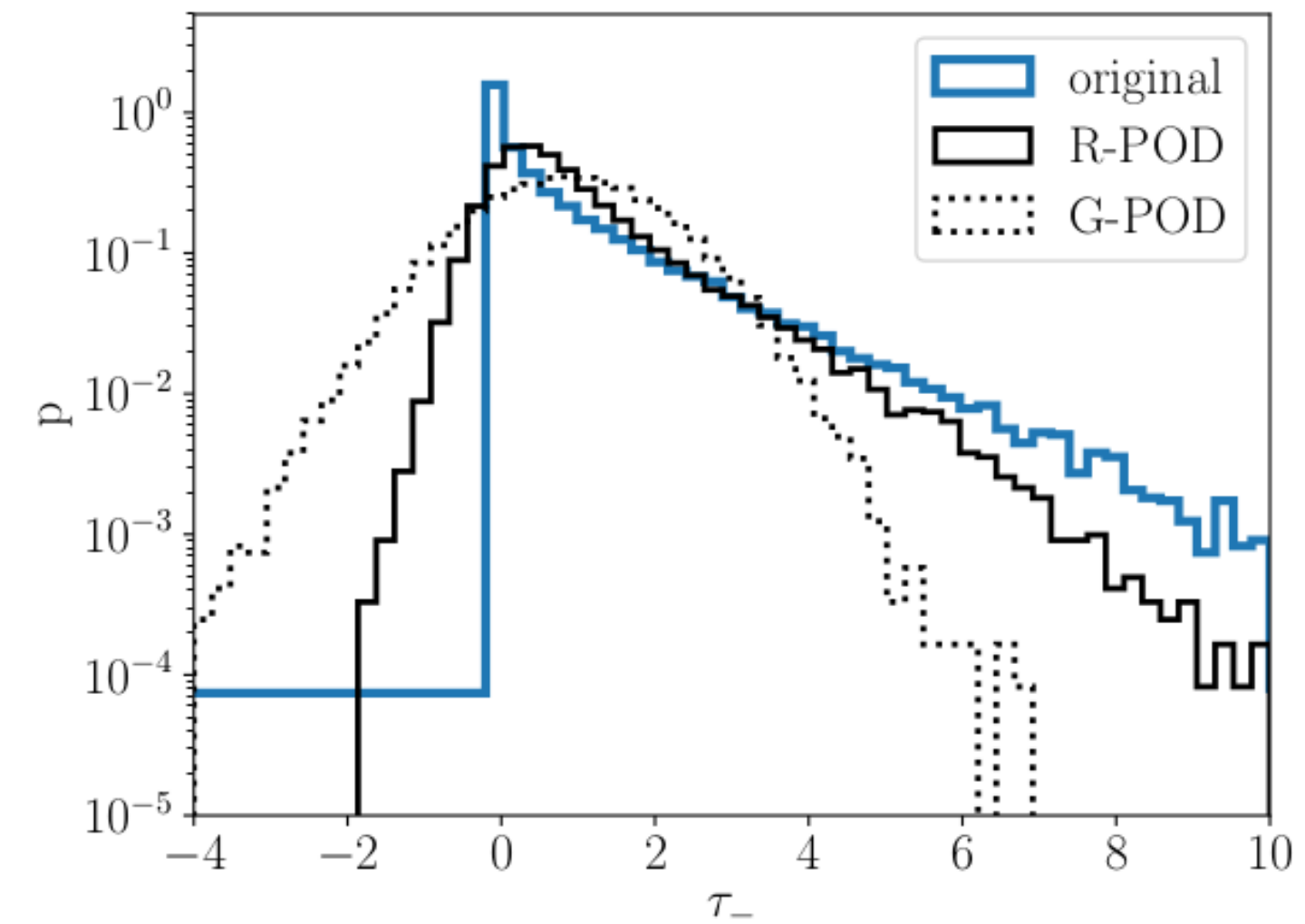
Distribution of instantaneous Reynolds stress τ_-



$$\tau_{-,jk}^{LDA} = \sum_n \theta_{kn} \varphi_{nj}$$

R-LDA: θ_{kn} stored

G-LDA: θ_{kn} sampled



$$\tau_{-,jk}^{POD} = \sum_n a_{nk} \phi_n(x_j)$$

R-POD: a_{kn} stored

G-POD: a_{kn} sampled

Summary

- Adaptation of a statistical analysis technique used in natural language processing to the identification of structures/motifs in turbulent flows using Q- Reynolds stress events
- Characteristics of motifs identified with the technique are consistent with classical hypotheses associated with the attached eddy model
- The generative model associated with LDA captures the intermittent properties of the Reynolds stress.

More about LDA...

- Companion talk: Identification of sea level pressure anomaly patterns using Latent Dirichlet Allocation, L. Fery, B. Dubrulle, B. Podvin, F. Pons & D. Faranda (tomorrow at 3:20pm, 2C)
- Coherent structure identification in turbulent channel flow using latent Dirichlet allocation, M Frihat, B Podvin, L Mathelin, Y Fraigneau, F Yvon, Journal of Fluid Mechanics 2021