WAVELET-POWERED NEURAL NETWORKS FOR TURBULENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

One of the fundamental driving phenomena for climate effects is fluid turbulence in geophysical flows. Modeling these flows and explaining its associated spatio-temporal phenomena are notoriously difficult tasks. Navier-Stokes (NS) equations describe all the details of the fluid motions, but require accounting for unfeasibly many degrees of freedom in the regime of developed turbulence. Model reduction and surrogate modeling of turbulence is a general methodology aiming to circumvent this curse of dimensionality. Originally driven by phenomenological considerations, multiple attempts to model-reduce NS equations got a new boost recently with Deep Learning (DL), trained on the ground truth data, e.g. extracted from high-fidelity Direct Numerical Simulations (DNS). However, early attempts of building NNs to model turbulence has also revealed its lack of interpretability as the most significant shortcoming. In this paper we address the key challenge of devising reduced but, at least partially, interpretable model. We take advantage of the balance between strong mathematical foundations and the physical interpretability of wavelet theory to build a spatio-temporally reduced dynamical map which fuses wavelet based spatial decomposition with spatio-temporal modeling based on Convolutional Long Short Term Memory (C-LSTM) architecture. It is shown that the wavelet-based NN makes progress in scaling to large flows, by reducing computational costs and GPU memory requirements.

1 CHALLENGE OF LEARNING SPATIO-TEMPORAL PHYSICS

Multitude of research problems in earth and climate sciences are exceptionally complex to study and model with existing analysis tools because of their high-dimensionality, with thousands-to-millions degrees of freedom exhibiting *spatio-temporal* dynamics, non-linearity and chaos. One of the most pertinent problems combining all these factors is fluid turbulence which occurs in geophysical flows and influences much of near and long term climate effects. In an era where vast quantities of turbulence data are generated for studying these applications, building practically usable, physics-driven reduced order surrogate models becomes extremely challenging and important. Recent surge in devising NN-based reduced models of turbulence Hennigh (2017); Wu et al. (2019); King et al. (2018) including significant efforts from the computer graphics community Wiewel et al. (2019); Werhahn et al. (2019); Xie et al. (2018); Um et al. (2018) for flow visualization by applying powerful, but application-agnostic Deep Learning (DL) techniques, such as Generative Adversarial Networks Goodfellow et al. (2014) and Convolutional LSTM (C-LSTM) Networks Xingjian et al. (2015) has provided valuable tools boosting research in this important field of physics. However, majority of approaches used in this emerging field are limited to analysis based on two dimensional spatial projections of the originally 3+1 dimension (three-dimensional space and another dimension in time) spatio-temporal data-sets. Some of the state-of-the art spatio-temporal NN modeling architectures, like C-LSTM, have significant memory costs thus resulting in a limited utility to practical, e.g. climate and geophysical, datasets. Our main focus in this paper is on making C-LSTM tractable for large scientific spatio-temporal datasets like fully developed turbulence, outlined in Appendix A.

2 EXISTING STRATEGY: AUTOENCODER 3D CONVOLUTIONAL LSTM

An approach to building reduced modeling of massive 3D spatio-temporal turbulence datasets is described in Mohan et al. (2019). The main spatio-temporal modeling block, 3D-C-LSTM, was

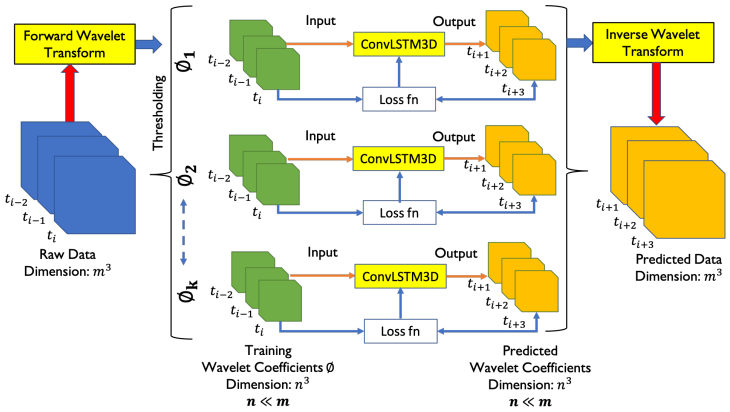


Figure 1: Wavelet-3D-C-LSTM: A schematic

implemented in Mohan et al. (2019) through 3D extension of the 2D C-LSTM, originally proposed in Xingjian et al. (2015). To reduce dimensionality, spatial compression and decompression steps were implemented via autoencoders, sandwiched by 3D-C-LSTM layers. A sequence of 50 temporal snapshots, each of size 128^3 with 3 velocity components, was used. This imposed a significant training cost, and the solution relied on convolutional auto-encoders to compress/decompress data before/after the 3D-C-LSTM block. The approach has helped to compress by factor of 125, down to $15^3 \times 15$ numbers. Knowledge of physics was utilized in Mohan et al. (2019) postfactum – only to evaluate prediction quality. However, this generally successful autoencoder-based approach had two important shortcomings. First, we do not have explicit control on the features to retain in the latent space and therefore some important features may be lost. Second, autoencoders are computationally expensive for even moderate in size datasets. This manuscript suggests to resolve these complications by replacing the autoencoder NNs with explicit and physics-based model reduction guided by wavelets. Wavelets provide additional benefits of strong mathematical foundations through the wavelet selection (e.g. resulting in numerical stability) and significant reduction of the underlying computational cost.

3 Proposed Solution: Wavelet-3D-C-LSTM

In this manuscript we propose a new NN scheme, coined *Wavelet-CLSTM*, to simultaneously address the twin challenges of reducing the computational cost and injecting physics-based features into the procedure. The key idea of our *Wavelet-CLSTM* scheme consists in decomposing the 3+1 dimensional training data set with the wavelet transform, which results in a compact representation though the wavelet coefficients. The approach is superior to the previously used (auto-encoder based compression/decompression) methodology because it represents turbulence data in a compact, mathematically accurate, robust and flexible way. Moreover, we capitalize on the fact, well documented in the literature Farge (1992); Meneveau (1991); Everson et al. (1990); Farge et al. (2001); Schneider et al. (1997); Pulido et al. (2016); Li et al. (2018), that the wavelet coefficients capture multi-scale physics embedded in the turbulence dataset in one of the most efficient compressed formats. For example, a 3 level wavelet decomposition of a volumetric dataset (Appendix A) of size 128^3 produces 512 coefficients of size 16^3. This reduction in dimensionality is critical for saving memory and improving practical applicability of powerful, but expensive, architectures like C-LSTM. Additionally, a full wavelet decomposition can be used to perfectly reconstruct the original dataset i.e it is non-lossy if all coefficients are considered. However, for reduced order modeling of practical datasets, we choose fewer coefficients to achieve maximal compression. This is an important choice to be made prior to training called **wavelet thresholding**, where the coefficients with the highest L2 norm are chosen for training - this cutoff number is determined by % of "energy" captured by the coefficients. A 3% thresholding would indicate $3/100 \times 512 \approx 15$ coefficients with the highest L2 norm. When compared with autoencoders, wavelets are advantageous because:

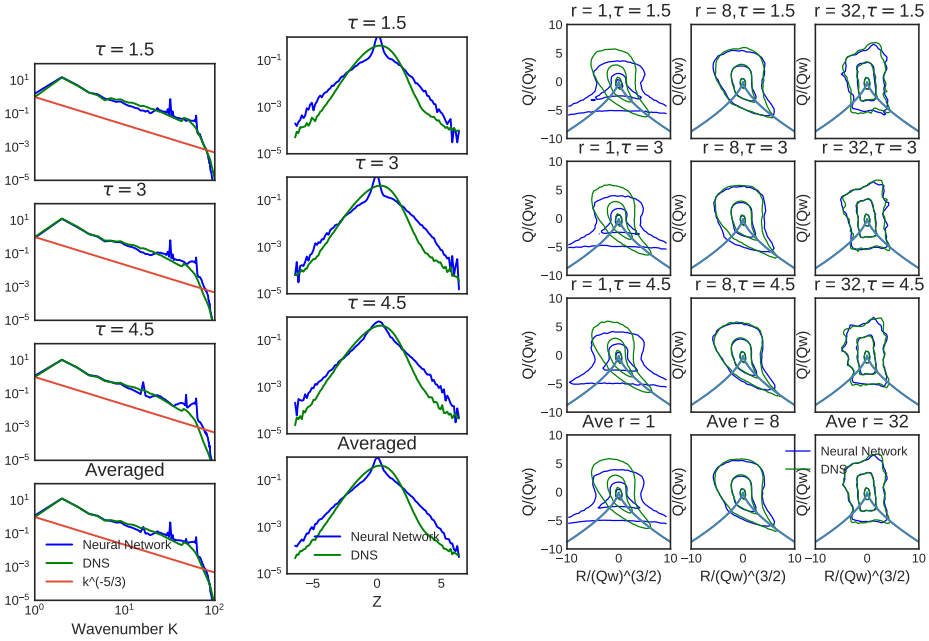
- Wavelet coefficients have a *dimensionality orders of magnitude lower than the original training data*, and low dimensional coefficients of a desired size can be computed for

extremely large datasets Pulido et al. (2016); Rodler (1999); Ihm & Park (1999). This decouples training for each wavelet coefficient from other coefficients, thereby avoiding communication overheads and memory limitations which otherwise would plague large, distributed, parallel training tasks.

- Wavelet transform (decomposition) and inverse wavelet transform (reconstruction) can be *computed analytically*, making it orders of magnitude cheaper and faster.
- An additional benefit of the analytical formulation translates into ability to extend thresholding, i.e. the scales to be modeled can be explicitly selected *a priori* to training. (This is to be contrasted to the convolutional autoencoder handicap which lacked direct control, apart from selection of kernel size Mohan et al. (2019) *during* training.)

Our *a priori* analysis of the 3% thresholding shows that it captures all the large scales, and a majority of the intermediate scales in turbulence, however excluding sufficiently small scales. This is an acceptable trade-off, since large and intermediate scales are typical quantities of interest in majority of practical applications Meneveau & Katz (2000); She & Leveque (1994); Porté-Agel et al. (2000). We would like to emphasize, however, that including smaller scales by increasing the thresholding percentage, is a degree of freedom to decide based on the application requirements. Increasing thresholding percentage increases the total training duration; but the adaptive, local nature of the coefficients ensures that the memory cost of training per coefficient stays *constant*, such that various coefficients can be trained *separately*, on available computer resources. This remarkable feature of the wavelet decomposition makes large scale parallelism a choice - rather than a necessity - thereby opening up this technique to extremely large datasets even with moderate computer resources available. To further increase compression efficiency we plan to investigate in the future scale based thresholding (i.e. different thresholds at different scales) as well as integer quantization (or re-quantization) to reduce the number of bits needed to represent the coefficients. A Schematic outlining this methodology is illustrated in Fig. 1.

4 RESULTS



(a) Energy Spectra.

(b) PDF of velocity gradients.

(c) Q-R plane: isolines of the velocity gradient invariants coarse-grained at small, $r = 1$, inertial, $r = 8$, and large, $r = 32$ scales.

Figure 2: *Wavelet-CLSTM* Neural Network Turbulence vs Physical Simulation (DNS)

The wavelet coefficients are computed with a *biorthogonal 1.3* Lee et al. (2019) mother wavelet and 3% thresholding which we only have 15 coefficients to train, out of a total of 512. We compare accuracy of the NN predictions based on the turbulence diagnostics developed and tested in King et al. (2018); Mohan et al. (2019). We predict a sequence of flow-fields from the trained model, and analyze the flow at $\tau = 1.5, 3$ and 4.5, which correspond to non-dimensional eddy turnover times. Analyzing the statistical properties of the predicted flow at varied time instants allows us to assess the long-term stability of our temporal predictions.

We now present results by applying 3 diagnostic tests of turbulence on the predicted flow. The diagnostics are explained in detail in Appendix B. First, we analyze relative significance of different HIT scales conducting the energy spectra test. Higher wave-numbers in Fig 2a correspond to smaller scales. It is clear from the results that the large scale spectra are matched almost exactly, with good reproduction in the intermediate scale range. Comparatively, small scale spectra are not reproduced well, which is intentional because a significant portion of small scales were removed (set to zero) during the thresholding. Effects of the small scale absence is also seen in the Probability Distribution Function (PDF) of the velocity gradient (Fig 2b), which tests solely the smallest scales of HIT. This is expected, since we are building a reduced order model for applications where large and inertial scales are of primary interest. The third test is the Q-R plane diagnostic in Fig 2c, which offers an arguably more stringent test of three-dimensional structure in turbulence Chertkov et al. (1999); Chong et al. (1990); Elsinga & Marusic (2010); Suman & Girimaji (2010), as described in the previous section. We observe in Fig. 2c that the *Wavelet-CLSTM* reproduces the large scale behavior almost perfectly, while reproduction of turbulence geometry start to deteriorate as we move down-scales, to intermediate ($r = 8$) and small ($r = 1$) scales. The small scale behavior is not reproduced due to the 3% thresholding favoring large scales. The symmetric structure seen in the small scale prediction is likely linked to the noise added by the model. The bottom graphic "Ave" in Fig. 2c shows the averaged diagnostics for the 3 time instants. Overall, the test results present ample evidence to the fact that due to the physically-interpretable selection of the wavelet basis, the *Wavelet-CLSTM* is capable of modeling the large and inertial scale spatio-temporal dynamics of HIT well. We point out that it is straightforward to include small scale behavior by including relevant wavelet coefficients, obviously on the expense of increase in the computational cost. Details of these diagnostic tests can be found in Appendix B.

5 Impact on Earth Sciences Research and Conclusion

Climate data is extremely high dimensional and necessitates model reduction for timely, efficient analysis and insights (Overpeck et al., 2011). Another major application is surrogate modeling of high-fidelity of geophysical flows (San & Maulik, 2018). We present here the first results for the novel *Wavelet-CLSTM*, which is an efficient, scalable, high dimensional deep NN framework for reduced modeling of turbulence, and similar or related multi-scale physical phenomena. The key strength of the framework is in the combination of a well-developed and mathematically justified wavelet decomposition with its highly desirable physical model reduction and interpretation power. Further investigation is desired into intelligent thresholding methods for non-stationary spatio-temporal climate phenomena. Another major application for *Wavelet-CLSTM* is for learning low dimensional dynamics of large climate and observational datasets where the governing equations are not well known, but where the wavelet coefficients can be a rigorous approach to identify and exploit multiscale patterns.

Acknowledgments

This work has been authored by employees of Triad National Security, LLC which operates Los Alamos National Laboratory (LANL) under Contract No. 89233218CNA000001 with the U.S. Department of Energy/National Nuclear Security Administration. Authors have been supported by LANL's LDRD program, project number 20190058DR. Author 1 also thanks the Center for Nonlinear Studies at LANL for support and acknowledges the ASC/LANL Darwin cluster for GPU computing infrastructure.

REFERENCES

Michael Chertkov, Alain Pumir, and Boris I Shraiman. Lagrangian tetrad dynamics and the phenomenology of turbulence. *Physics of fluids*, 11(8):2394–2410, 1999.

Min S Chong, Anthony E Perry, and Brian J Cantwell. A general classification of three-dimensional flow fields. *Physics of Fluids A: Fluid Dynamics*, 2(5):765–777, 1990.

Don Daniel, Daniel Livescu, and Jaiyoung Ryu. Reaction analogy based forcing for incompressible scalar turbulence. *Physical Review Fluids*, 3(9):094602, 2018.

GE Elsinga and I Marusic. Universal aspects of small-scale motions in turbulence. *Journal of Fluid Mechanics*, 662:514–539, 2010.

R Everson, L Sirovich, and KR Sreenivasan. Wavelet analysis of the turbulent jet. *Physics Letters A*, 145(6-7):314–322, 1990.

Marie Farge. Wavelet transforms and their applications to turbulence. *Annual review of fluid mechanics*, 24(1):395–458, 1992.

Marie Farge, Giulio Pellegrino, and Kai Schneider. Coherent vortex extraction in 3d turbulent flows using orthogonal wavelets. *Physical Review Letters*, 87(5):054501, 2001.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.

Oliver Hennigh. Lat-net: compressing lattice boltzmann flow simulations using deep neural networks. *arXiv preprint arXiv:1705.09036*, 2017.

Insung Ihm and Sanghun Park. Wavelet-based 3d compression scheme for interactive visualization of very large volume data. In *Computer graphics forum*, volume 18, pp. 3–15. Wiley Online Library, 1999.

Ryan King, Oliver Hennigh, Arvind Mohan, and Michael Chertkov. From deep to physics-informed learning of turbulence: Diagnostics. *arXiv preprint arXiv:1810.07785*, 2018.

Gregory R Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O'Leary. Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019.

Shaomeng Li, Nicole Marsaglia, Christoph Garth, Jonathan Woodring, John Clyne, and Hank Childs. Data reduction techniques for simulation, visualization and data analysis. In *Computer Graphics Forum*, volume 37, pp. 422–447. Wiley Online Library, 2018.

D Livescu, J Mohd-Yusof, MR Petersen, and JW Grove. Cfdns: a computer code for direct numerical simulation of turbulent flows. *Los Alamos National Laboratory Technical Report No. LA-CC-09-100*, 2009.

Charles Meneveau. Analysis of turbulence in the orthonormal wavelet representation. *Journal of Fluid Mechanics*, 232:469–520, 1991.

Charles Meneveau and Joseph Katz. Scale-invariance and turbulence models for large-eddy simulation. *Annual Review of Fluid Mechanics*, 32(1):1–32, 2000.

Arvind Mohan, Don Daniel, Michael Chertkov, and Daniel Livescu. Compressed convolutional lstm: An efficient deep learning framework to model high fidelity 3d turbulence. *arXiv preprint arXiv:1903.00033*, 2019.

Jonathan T Overpeck, Gerald A Meehl, Sandrine Bony, and David R Easterling. Climate data challenges in the 21st century. *science*, 331(6018):700–702, 2011.

Fernando Porté-Agel, Charles Meneveau, and Marc B Parlange. A scale-dependent dynamic model for large-eddy simulation: application to a neutral atmospheric boundary layer. *Journal of Fluid Mechanics*, 415:261–284, 2000.

Jesus Pulido, Daniel Livescu, Jonathan Woodring, James Ahrens, and Bernd Hamann. Survey and analysis of multiresolution methods for turbulence data. *Computers & Fluids*, 125:39–58, 2016.

Flemming Friche Rodler. Wavelet based 3d compression with fast random access for very large volume data. In *Proceedings. Seventh Pacific Conference on Computer Graphics and Applications (Cat. No. PR00293)*, pp. 108–117. IEEE, 1999.

Omer San and Romit Maulik. Extreme learning machine for reduced order modeling of turbulent geophysical flows. *Physical Review E*, 97(4):042322, 2018.

Kai Schneider, NK-R Kevlahan, and Marie Farge. Comparison of an adaptive wavelet method and nonlinearly filtered pseudospectral methods for two-dimensional turbulence. *Theoretical and computational fluid dynamics*, 9(3-4):191–206, 1997.

Zhen-Su She and Emmanuel Leveque. Universal scaling laws in fully developed turbulence. *Physical review letters*, 72(3):336, 1994.

Sawan Suman and Sharath S Girimaji. Velocity gradient invariants and local flow-field topology in compressible turbulence. *Journal of Turbulence*, (11):N2, 2010.

Kiwon Um, Xiangyu Hu, and Nils Thuerey. Liquid splash modeling with neural networks. In *Computer Graphics Forum*, volume 37, pp. 171–182. Wiley Online Library, 2018.

Maximilian Werhahn, You Xie, Mengyu Chu, and Nils Thuerey. A multi-pass gan for fluid flow super-resolution. *arXiv preprint arXiv:1906.01689*, 2019.

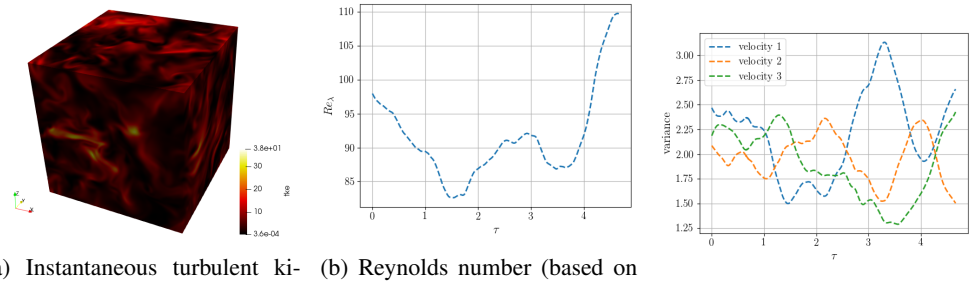
Steffen Wiewel, Moritz Becher, and Nils Thuerey. Latent space physics: Towards learning the temporal evolution of fluid flow. In *Computer Graphics Forum*, volume 38, pp. 71–82. Wiley Online Library, 2019.

Jin-Long Wu, Karthik Kashinath, Adrian Albert, Dragos Chirila, Heng Xiao, et al. Enforcing statistical constraints in generative adversarial networks for modeling chaotic dynamical systems. *arXiv preprint arXiv:1905.06841*, 2019.

You Xie, Erik Franz, Mengyu Chu, and Nils Thuerey. tempogan: A temporally coherent, volumetric gan for super-resolution fluid flow. *ACM Transactions on Graphics (TOG)*, 37(4):95, 2018.

SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pp. 802–810, 2015.

A 3D Homogeneous Isotropic Turbulence Dataset and Training



(a) Instantaneous turbulent kinetic energy

(b) Reynolds number (based on Taylor microscale)

(c) Individual velocity variances

Figure 3: Representative Statistics of the Simulation

The dataset consists of a 3D Direct Numerical Simulation (DNS) of homogeneous, isotropic turbulence, in a box of size 128^3. We denote this dataset as *HIT* for the remainder of this work. We provide a brief overview of the simulation and its physics in this section, and a detailed discussion

can be found in Daniel et al. (2018). The ScalarHIT dataset is obtained using the incompressible version of the CFDNS (Livescu et al., 2009) code, which uses a classical pseudo-spectral algorithm. We solve the incompressible Navier-Stokes equations:

$$\partial_{x_i} v_i = 0, \qquad \partial_t v_i + v_j \partial_{x_j} v_i = -\frac{1}{\rho}\partial_{x_i} p + \nu \Delta v_i + f_i^v,$$

where f^v is a low band forcing, restricted to small wavenumbers $k < 1.5$ [1]. The 128^3 pseudo-spectral simulations are dealiased using a combination of phase-shifting and truncation to achieve a maximum resolved wavenumber of $k_{max} = \sqrt{2}/3 \times 128 \sim 60$.

For illustration, Figure 3a shows the turbulent kinetic energy at a time instant. Figure 3b shows the variation in the Taylor-microscale based Reynolds number with the eddy turnover time, which characterizes the large turbulence scales. Finally, the variances in all 3 velocity components are shown in Fig. 3c. Based on the sampling rate, each eddy turnover time τ consists of 33 snapshots. The training dataset uses 22 snapshots $\approx 0 - 0.75\tau$ and test dataset also consists of 22 snapshots in $\approx 4 - 4.75\tau$.

B Some Diagnostic Tests of Turbulence

We now briefly describe 3 basic tests of 3D turbulence which are used as "diagnostic" metrics in this work, for the accuracy of the flow predicted by the trained model.

B.1 $4/5$ Kolmogorov law and the Energy Spectra

The main statement of the Kolmogorov theory of turbulence is that asymptotically in the inertial range, i.e. at $L \gg r \gg \eta$, where L is the largest, so-called energy-containing scale of turbulence and η is the smallest scale of turbulence, so-called Kolmogorov (viscous) scale, $F(r)$ does not depend on r. Moreover, the so-called $4/5$-law states for the third-order moment of the longitudinal velocity increment

$$L \gg r \gg \eta: \quad S_3^{(i,j,k)} \frac{r^i r^j r^k}{r^3} = -\frac{4}{5}\varepsilon r, \tag{1}$$

where $\varepsilon = \nu D_2^{(i,j;i,j)}/2$ is the kinetic energy dissipation also equal to the energy flux.

Self-similarity hypothesis extended from the third moment to the second moment results in the expectation that within the inertial range, $L \gg r\eta$, the second moment of velocity increment scales as, $S_2(r) \sim v_L(r/L)^{2/3}$. This feature is typically tested by plotting the energy spectra of turbulence (expressed via $S_2(r)$) in the wave vector domain, e.g. as shown in the results section.

B.2 Intermittency of Velocity Gradient

Consequently from Eqn. 1, the estimation of the moments of the velocity gradient results in

$$D_n \sim \frac{S_n(\eta)}{\eta^n}. \tag{2}$$

This relation is strongly affected by intermittency for large values of n (i.e. extreme non-Gaussian behavior) of turbulence, and is a valuable test of small scale behavior.

B.3 Statistics of coarse-grained velocity gradients: $Q - R$ plane.

Both the diagnostics described so far are highly averaged quantities which do not explicitly account for the flow structure and 3D effects of the predicted velocities. To address this, we utilize isolines of probability in the $Q - R$ plane, which expresses intimate features of the turbulent flow topology, having a nontrivial shape documented in literature. We compute two invariants Q and R from the

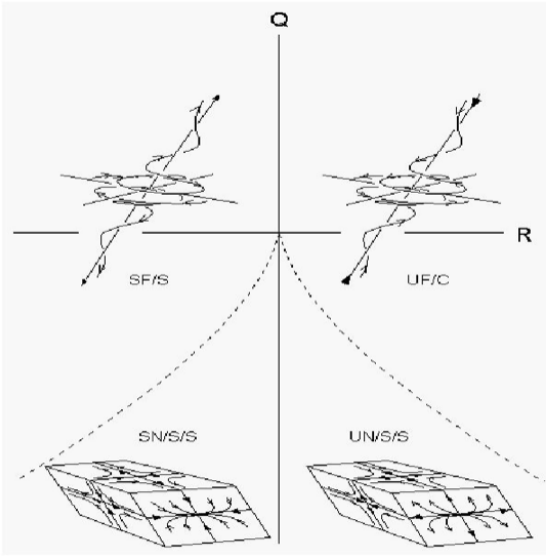


Figure 4: Analysis of Stretching and compression of turbulent structures via Q-R plane PDFs

gradient of the velocity tensor V, as shown in Eqn. 5

$$M = (\nabla V)_r \tag{3}$$

$$R = \frac{-tr(M^3)}{3} \tag{4}$$

$$Q = \frac{-tr(M^2)}{2} \tag{5}$$

See Chertkov et al. (1999) and references therein. Different parts of the $Q - R$ plane are associated with different structures of the flow, as illustrated in Fig. 4. Thus, lower right corner (negative Q and R), which has higher probability than other quadrants, corresponds to a pancake type of structure (two expanding directions, one contracting) with the direction of rotation (vorticity) aligned with the second eigenvector of the stress. This tear-drop shape of the probability isoline becomes more prominent with decrease of the observation scale r. Here, we study the $Q - R$ plane filtered at different scales r, to examine large ($r = 32$), inertial ($r = 8$), and small scale ($r = 1$) behaviors. This allows us to selectively analyze the accuracy of our predictions at different scales, since we are interested in primarily the large and inertial ranges for reduced order modeling.