

Using autoencoders to enhance data-driven turbulence modeling in squircular geometries

By M. S. S. Macedo[†], B. P. Brener[†], R. L. Thompson[†], T. Zahtila, S. Mirjalili, M. P. Whitmore AND G. Iaccarino

This work uses data from direct numerical simulations (DNS) of internal flows in geometries that are related by a squareness transform to examine the potential for autoencoders to compress data sets of turbulent mean fields into low-order representations. Continuous transformation from a circle cross section to a square results in intermediate squircle geometries. Starting from the canonical doubly periodic internal flows, such as channels or pipes, only one component of velocity is present in the mean field. As the geometry becomes more square, secondary flows are intensified, and therefore, accurate compression of the mean flow quantities becomes a distinct challenge. Compression of the mean fields is supplemented by analysis that leads to functional forms of the latent-space parameters. Subsequently, the latent space functional forms were employed to synthetically generate unseen cases and verified as predictive.

1. Introduction

Wall-bounded flow computations have generated many landmark studies, of pertinence to the present effort, in pipes (Wu & Moin 2008) and ducts (Pirozzoli *et al.* 2018). Pipes attract attention firstly because of their engineering ubiquity and because their homogeneous streamwise- and azimuthal-flow directions facilitate statistical analysis. Ducts with complex cross-sections have attracted research interest since the seminal experiments of Nikuradse (1930), owing to the secondary motions that can be observed, which are not present in pipes. The Fernández-Guasti squircle is a smooth algebraic curve that maps between a circle and square, parameterized by the squareness, s , from 0 (circle) to 1 (square). The squircle of unit minor radius is given by the quartic polynomial $x^2 + y^2 - sx^2y^2 = 1$ (Guasti 1992). Starting from a disc of points, the transformation to a square (Fong 2016) is given by,

$$\begin{aligned} x &= \frac{\text{sgn}(uv)}{v\sqrt{2}} \cdot \sqrt{\frac{u^2 + v^2 - \sqrt{(u^2 + v^2)(u^2 + v^2 - s \cdot 4.0 \cdot u^2v^2)}}{s}}, \\ y &= \frac{\text{sgn}(uv)}{u\sqrt{2}} \cdot \sqrt{\frac{u^2 + v^2 - \sqrt{(u^2 + v^2)(u^2 + v^2 - s \cdot 4.0 \cdot u^2v^2)}}{s}}, \end{aligned} \tag{1.1}$$

where u and v are the original distribution of points on the closed disc. We anticipate that as the squareness factor s of the geometry is increased, the intensity of the arising secondary flows will increase.

[†] Department of Mechanical Engineering, Federal University of Rio de Janeiro, Brazil

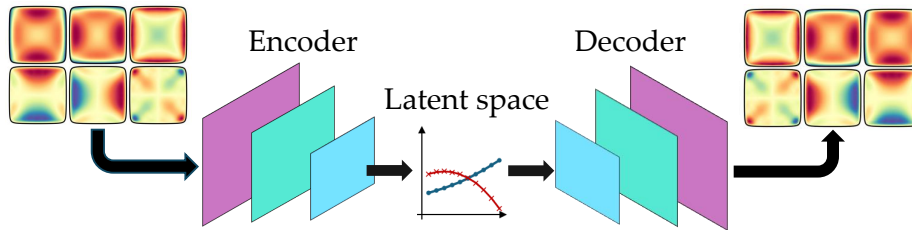


FIGURE 1. Illustration of the autoencoder architecture

1.1. Autoencoders

Autoencoders are a powerful machine learning technique capable of performing extreme data compression with minimal losses. The architecture is segregated in an encoder, which compresses the data, and a decoder, which decompresses it back to the original dimensions. The process is done incrementally by the consecutive layers of the two segments. The output of the encoder, where data is at the maximum compression stage, is denoted the latent space. An illustration of the autoencoder architecture is displayed in Figure 1.

This capability has caught the attention of many authors in the turbulence and fluid dynamics community. An appealing application is to use it in reduced order models (ROM), where a typical approach is to use proper orthogonal decomposition (POD) to reduce dimensionality. Multiple works have demonstrated that the inherent nonlinearity of the latent representations significantly outperforms POD. In Murata *et al.* (2020), the authors have verified that a single latent variable contained contributions from multiple POD modes. Recently, many data-driven ROM works have explored this approach [e.g., Lee & Carlberg (2019); Agostini (2020); Fukami *et al.* (2020); Mohan *et al.* (2020); Zhang (2023); Racca *et al.* (2023); Wang *et al.* (2024)].

Embedding the latent space with physical meaning was one of the concerns in Saetta *et al.* (2023). Autoencoders were used to analyze a data set with Reynolds-averaged Navier-Stokes (RANS) simulations of airfoils. Clear connections between the low-dimensional latent space and the flow parameters, such as the angle of attack, were observed. These findings enabled the prediction of aerodynamic stall and the manipulation of the latent space to predict flow cases unseen in training. In Brener *et al.* (2024b), the current authors investigated the latent representations of DNS data sets of pressure-driven wall-bounded flows. In the investigated plane channel and square duct flows, the latent space and the friction and bulk Reynolds numbers were closely correlated. Depending on the training configuration, the latent space was a direct function of the friction factor. As in Saetta *et al.* (2023), the functional forms established with the latent space were used to predict unseen flow cases.

Following Brener *et al.* (2024b), in this work we have trained autoencoders with the novel squircular duct data set. This data set differs from the conventional plane channel and square duct flows mainly in its parameterization not only by the Reynolds number but also by the geometry parameter s . With the objective of predicting unseen cases, we interpret the latent space variables in light of the physical quantities. Contrary to Saetta *et al.* (2023) and Brener *et al.* (2024b), the principal quantity of interest here is the Reynolds stress tensor (RST). With the latent space predictions of the RST we investigate whether autoencoders are suitable to provide data-driven closure for the RANS equations. This is related to the many efforts in the literature of modeling turbulence with machine learning.

Conventionally, other techniques have been employed in the data-driven modeling literature, such as vanilla neural networks (Ling *et al.* 2016; Cruz *et al.* 2019; Berrone & Oberto 2022; Macedo *et al.* 2024) and random forests (Wang *et al.* 2017; Wu *et al.* 2018; Kaandorp & Dwight 2020; Brener *et al.* 2024a). Although the RST is the immediate closure quantity in RANS modeling, it has been demonstrated in Thompson *et al.* (2016), Andrade *et al.* (2018), Wu *et al.* (2019) and Brener *et al.* (2021) that using it in data-driven efforts is tied to significant accuracy losses. Despite that, as a first step, we have worked with the RST and in future work we intend to extend this methodology to the prediction of other quantities, such as the Reynolds force vector (RFV) of Cruz *et al.* (2019) and variations of the RST (Wu *et al.* 2018) and the RFV (Brener *et al.* 2021).

2. Methodology

2.1. Squire direct numerical simulation database

We utilize DNS to simulate statistically steady flow through a squire. The code used in this study is charLES, which is a low-order, finite-volume code with low-dissipation numerical methods (Bres *et al.* 2018), and presently, the low-Mach formulation is used. The code uses meshes based on centroidal Voronoi tessellations generated from hexagonally close-packed seed points (Whitmore *et al.* 2024). The seed points for grid refinement are generated via an O-grid mesh for the circle ($s = 0$) and all other cross sections other than the square ($s = 1.0$). For the square limit, the near-wall region was refined in accordance with a hyperbolic stretching function.

In order to maintain dynamic similarity across the simulation sets, the pressure gradient P that drives the flow was adjusted between the cases to achieve a constant friction velocity u_τ . Therefore, the friction Reynolds number Re_τ , based off of the minor axis of the geometry, viscosity ν and friction velocity u_τ , is held constant across geometries. However, the hydraulic radius will vary and is not constrained, which forms part of the challenge for the modeling of the present data set.

As part of the present campaign, eight cases were run in total. This data set contains five simulations conducted at friction Reynolds number, $Re_\tau = 180$, with squareness transform parameter, $s = \{0.0, 0.6, 0.85, 0.97, 1.0\}$. Note, this unequally spaced choice of squire transform parameter is chosen to generate equal spacing of the squire major axis. Thereafter, three cases were run with the pairs of friction Reynolds number Re_τ and squareness transform parameter s of (540, 0.0), (360, 0.85), and (540, 1.0). With computational cost allocation in mind, this forms sparse sampling of the $Re_\tau - s$ space under investigation. The seed points that are input to the Voronoi-tessellated computational mesh for the friction Reynolds number $Re_\tau = 180$ case are stretched in the wall-normal direction, and varied from near-wall to centerline with $\Delta y^+ = [0.5, 2.5]$, in the streamwise direction, $\Delta x^+ = 5.14$, and in the azimuthal direction, and $R\Delta\theta^+ \approx 3.1$, with slight variance according to the squareness parameter. In the case of the duct, the cross-stream directional spacing Δz^+ is the same as that of the wall-normal stretching. For the $Re_\tau = 540$ cases, grid spacing was slightly relaxed to $\Delta x^+ = 6.17$ and $\Delta y^+ = [0.7, 3.3]$ due to their significantly increased computational cost. The streamwise length of the domain was set to $L_x = 4\pi R$, which is sufficient to capture energized structures present in an internal wall flow calculation (Chin *et al.* 2010) and provide accurate flow statistics of interest here. In line with Vinuesa *et al.* (2016), to account for the impact of the computational domain's streamwise length on statistical convergence, we use effective time averaging intervals, $\Delta t^* = \Delta t_{av} L_x / 6h$, expressed in terms of eddy turnover time, h/u_τ .

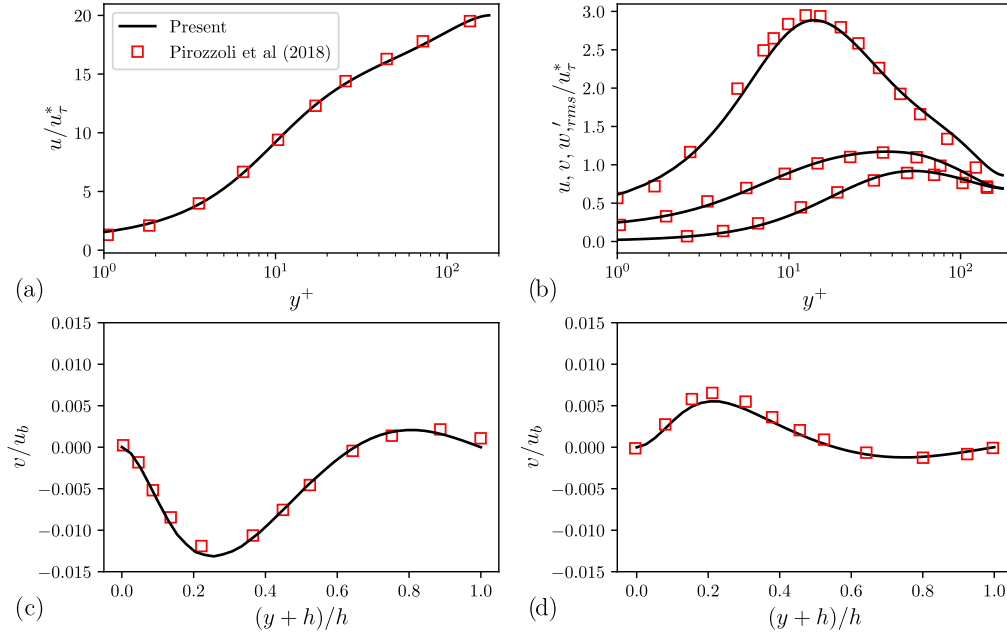


FIGURE 2. Profiles of (a) mean streamwise velocity \bar{u}/u_τ^* along the wall bi-sector. (b) Normal turbulence intensities and cross-stream velocity at (c) $z/h = -0.75$ and (d) $z/h = -0.25$. The data of Pirozzoli *et al.* (2018) are displayed for comparison. Here, u_τ^* is the friction velocity averaged across the entire duct, and u_b is the bulk velocity.

The effective time-averaging interval for the $Re_\tau = 180$ cases was $\Delta t^* = 1256$, which is comparable to the preceding duct computations that have been carried out (Vinuesa *et al.* 2014; Pinelli *et al.* 2010). For the $Re_\tau = 360$ and 540 cases, a restricted time interval of $\Delta t^* = 420$ was set for the collection of statistics.

In Figure 2, validation of flow quantities for the duct-limiting case against high-quality reference work is presented, with good agreement observed. This case was chosen because of the strong secondary motions present and so forms a stringent validation case compared to the remaining geometries under investigation.

2.2. Autoencoder architecture and training

A convolutional autoencoder was employed to analyze the squircle duct DNS data set. With an architecture similar to that used in Brener *et al.* (2024b), they were built and trained using the TensorFlow (Abadi *et al.* 2015) Python library. The architecture is summarized in Table 1, where the encoder layers are listed. The decoder layers were suppressed, as they are same as the encoder ones, but arranged in the reverse order. Average pooling layers were used to reduce dimensions, while upsampling layers were used to increase them in the decoder. The convolutional layers used 9×9 filters and the ReLU activation function; in the output layer the sigmoid activation function was used to ensure that the outputs were within the range $[0; 1]$. The Adam optimizer (Kingma & Ba 2015) was used with the mean squared error as the loss function and a training interruption criteria of 10^{-5} . An adjustable learning rate departing from the initial value of 10^{-3} was employed, and it was progressively decreased until a final value of 10^{-4} .

To compare the latent space when training with the velocity field or the RST, we have trained autoencoders for both quantities. The components of the quantities (three

TABLE 1. Encoder architecture. P is the dimension of the latent space and N is the number of components of the input field, i.e., three for the velocity and six for the Reynolds stress.

| <i>Layer type</i> | <i>Filters</i> | <i>Pooling size</i> | <i>Output shape</i> |
|-------------------|----------------|---------------------|--------------------------|
| Input | - | - | $64 \times 64 \times N$ |
| Convolutional | 32 | - | $64 \times 64 \times 32$ |
| Average pooling | - | 2×2 | $32 \times 32 \times 32$ |
| Convolutional | 16 | - | $32 \times 32 \times 16$ |
| Average pooling | - | 4×4 | $8 \times 8 \times 16$ |
| Convolutional | 4 | - | $8 \times 8 \times 4$ |
| Average pooling | - | 4×4 | $2 \times 2 \times 4$ |
| Convolutional | P | - | $2 \times 2 \times P$ |
| Average pooling | - | 2×2 | P |

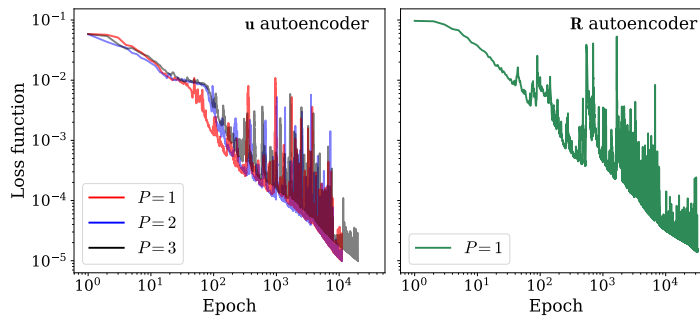


FIGURE 3. Temporal evolution of the training loss for the velocity \mathbf{u} and Reynolds stress \mathbf{R} autoencoders. P is the dimension of the latent space.

for \mathbf{u} and six for \mathbf{R}) were arranged as channels of the inputs and outputs. Prior to training, the data sets were normalized to be within the range $[0; 1]$, using the expression of Eq. 2.1 below, where q_i is the i component of the quantity of interest (\mathbf{u} or \mathbf{R}), q_i^* is the normalized quantity, and $\min(Q_i)$ and $\max(Q_i)$ denote the minimum and maximum values of q_i across all simulations in the data set,

$$q_i^* = \frac{q_i - \min(Q_i)}{\max(Q_i) - \min(Q_i)}. \quad (2.1)$$

In the present work we have still only trained the autoencoders with the five geometries of the same $Re_\tau = 180$. As the training data were only parameterized by geometry, we used a latent space dimension of $P = 1$. Increasing the dimensionality did not provide gains in training or in the accuracy of the outputs. Figure 3 demonstrates the evolution of training for the velocity and RST autoencoders. The temporal evolution with $P = 2$ and $P = 3$ was also presented for \mathbf{u} . As observed, training was not benefited by the dimensional increase; in fact, the autoencoder took longer to converge in the $P = 3$ case.

An important aspect of training with different geometries is defining how to present the data to the autoencoder. Two strategies were evaluated to sample the data from the charLES mesh. In the first strategy, the data were sampled in 90 azimuthal and 200 radial positions in all simulations. Although this strategy guaranteed the same number of points for all geometries, once they were arranged in square matrices, the spatial

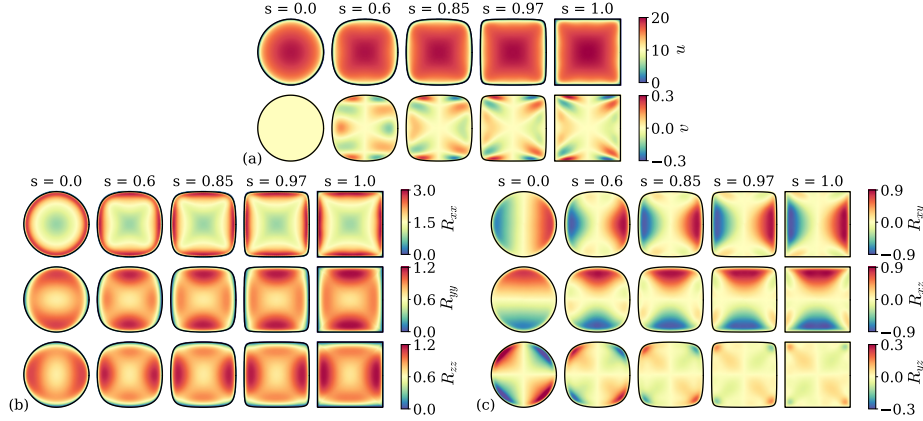


FIGURE 4. Time-averaged DNS quantities: (a) velocity components u and v . (b) Normal and (c) shear components of the Reynolds stress \mathbf{R} , with $Re_\tau = 180$.

coordinates of the data at a given position i, j for one geometry were not the same for the same position but came from the other geometry matrix. In the second strategy, an evenly spaced square grid was used to sample the points from the geometries. For the non-square cases, the points outside the duct walls were padded with zeros. The second strategy was adopted because its results were better in terms of both the accuracy of the autoencoder outputs and the behavior of the latent space. In this manner, the velocities and the RST of each case were interpolated to a 128×128 square mesh. The autoencoders were trained with a single averaged quadrant of the cross sections, resulting in the input and output dimensions of 64×64 . Other sampling strategies will be addressed in our future work.

3. Results

In this section we first present the averaged DNS velocities and RST and then present the autoencoder results. Because the autoencoders were trained only with averaged quantities, from this point onward, for simplicity, the overline denoting the time averaging has been dropped. Therefore, for the remainder of this work u and v denote the streamwise and cross-stream mean velocities, respectively.

The averaged velocities u and v of the five simulated geometries are presented in Figure 4(a) for the $Re_\tau = 180$ case. For brevity, the component w was not shown, as it is identical to v but rotated with respect to the duct's center. The strengthening of the recirculation is clearly noticed as the squareness parameter s is increased. The components of the Reynolds stress \mathbf{R} are presented in Figure 4(b,c) for the same $Re_\tau = 180$.

3.1. Autoencoder results

To demonstrate the effectiveness of the proposed autoencoder architecture in compressing the data to a single variable, we present in Figure 5 its outputs when training with the velocities of the five geometries, with $Re_\tau = 180$. For brevity, only the components u and v for three geometries are presented. The reconstructions of the other geometries, and also of the RST autoencoder, were in qualitative agreement with those of Figure 5. The zero padding imposed onto the non-square geometries is promptly observed. The errors

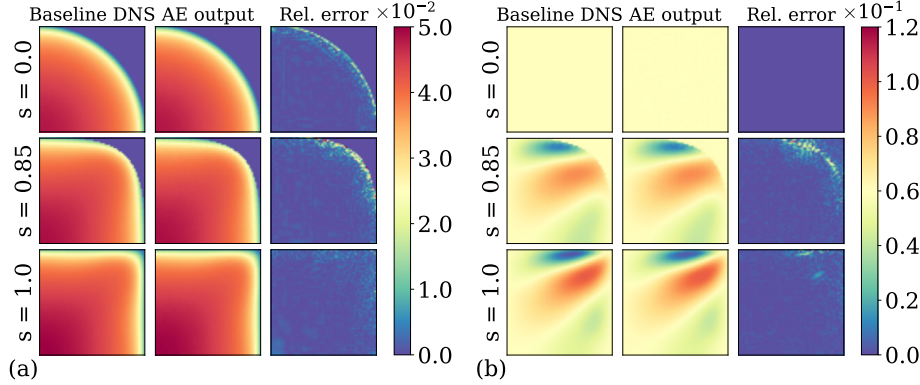


FIGURE 5. (a) Primary and (b) secondary velocity components reconstructed by the autoencoder. The colorbar is only representative of the relative errors. For conciseness, only three geometries are presented.

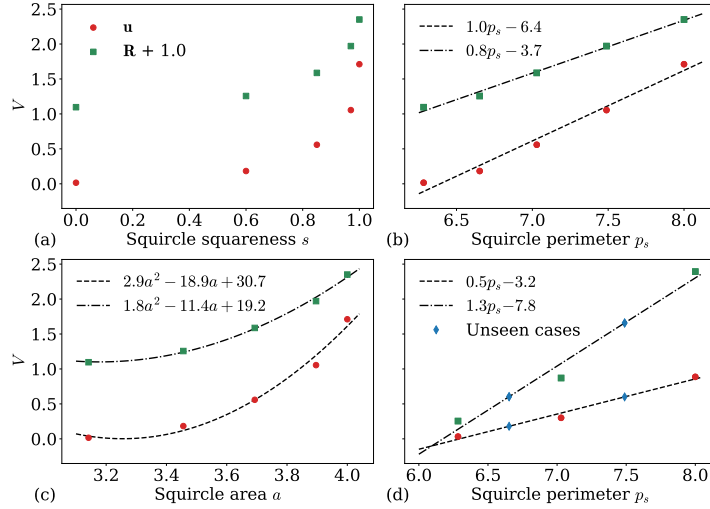


FIGURE 6. Velocity and Reynolds stress latent spaces as functions of (a) the squareness s , (b) the perimeter p_s and (c) the cross section area a . (d) Latent space trained with the geometries $s = 0.0$, $s = 0.85$ and $s = 1.0$ as a function of the perimeter p_s .

in the last column of Figures 5(a,b) are relative to the maximum absolute value of each component and demonstrate that the autoencoder has most difficulty in predicting the wall contours.

The latent variable V of the velocity and RST autoencoder was analyzed with respect to three geometrical parameters: the squareness s , the perimeter p_s and the cross section area a . Although the immediate choice was s , in Figure 6(a) it is observed that establishing a functional relation of the form $V(s)$ was not trivial. On the other hand, the latent variable V could be easily expressed as simple functions of p_s and a . In Figure 6(a) linear correlations with expressions of $V = 1.01p_s - 6.45$ and $V = 0.76p_s - 3.72$ for the velocity and RST were observed, respectively. With respect to the area a , the latent variable followed an second order polynomial, with expressions of $V = 2.9a^2 - 18.87a + 30.71$ for the velocity and $V = 1.79a^2 - 11.38a + 19.17$ for the RST.

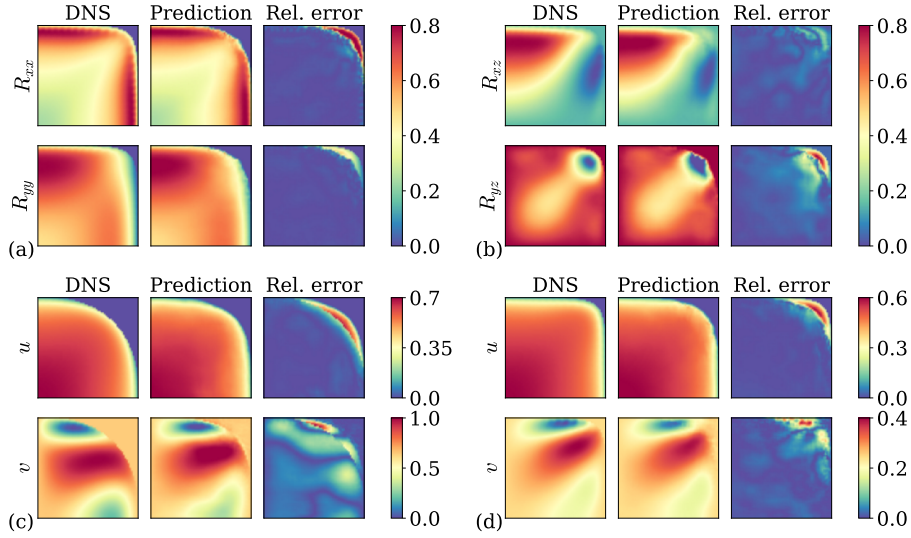


FIGURE 7. (a) Normal and (b) shear components of the Reynolds stress predicted by the decoder for the $s = 0.97$ case. Predicted velocities for the (c) $s = 0.60$ and (d) $s = 0.97$ cases.

Given that $V(p_s)$ behaves linearly, we have proceeded to use this function to obtain predictions of cases unseen in training. As a first attempt, the velocity and RST autoencoders were trained with the three geometries $s = 0.0$, $s = 0.85$ and $s = 1.0$. Using the three latent variables of the training cases, a linear fit with respect to p_S was computed for each quantity and the unseen cases of $s = 0.60$ and $s = 0.97$ were predicted by the decoder. Figure 6(d) presents the training and predicted cases as well as the fitted latent function.

Two normal and two shear components of the predicted RST for the $s = 0.97$ case are presented in Figure 7(a,b). The remaining components for the $s = 0.97$ and $s = 0.60$ cases were omitted for brevity but were qualitatively similar to those in Figure 7. Once again, the relative errors with respect to the maximum of the components reveal the highest errors in the squircle wall. This particularly impairs the prediction of the R_{yz} component, as an important segment of it is located close to the duct's corner. Figure 7(c,d) presents the mean velocity decoder predictions for both unseen cases $s = 0.60$ and $s = 0.97$, with the highest deviation again at the walls. For the $s = 0.60$ case, the prediction is more square than expected, whereas for the $s = 0.97$ case, the prediction is more rounded.

4. Conclusions

A novel DNS data set of pressure-driven wall-bounded turbulent flows was computed using the charLES finite-volume software. The squircle geometry was used as the cross section area of ducts with a total of five shapes, including the circle and the square. The strengthening of the secondary motion was observed as the squareness s was increased. Besides the geometry, the data set is also parameterized by different Re_τ .

The squircle data set was tailored to be used by machine learning in data-driven turbulence modeling. With this purpose, we have employed autoencoders to compress the simulations to a unitary latent space. It was observed that the latent space was a linear function of the squircle perimeter and a quadratic function of the area. The simple

functional relationships with the squircle geometry demonstrate that an interpretable compression was achieved. This important finding enabled the manipulation of the linear function to predict unseen cases, indicating that the autoencoders may be used as a data-driven turbulence model.

Future stages of this research will address the propagation of the predicted RST in the RANS equations, as well as evaluate other closures. Improving the accuracy of predictions, particularly the wall contours, will be paramount in that effort. Detailed investigations of the effects of the autoencoder architecture, data treatment, different training strategies and using combinations of both geometry and Reynolds numbers will also be the object of our future work.

REFERENCES

- ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G. S., DAVIS, A., DEAN, J., DEVIN, M. *et al.* 2015 TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
- AGOSTINI, L. 2020 Exploration and prediction of fluid dynamical systems using autoencoder technology. *Phys. Fluids* **32**, 067103.
- ANDRADE, J. R., MARTINS, R. S., THOMPSON, R. L., MOMPEAN, G. & NETO, A. S. 2018 Analysis of uncertainties and convergence of turbulent wall-bounded flows by means of a physically-based criterion. *Phys. Fluids* **30**, 045106.
- BERRONE, S. & OBERTO, D. 2022 An invariances-preserving vector basis neural network for the closure of Reynolds-averaged Navier–Stokes equations by the divergence of the Reynolds stress tensor. *Phys. Fluids* **34**, 095136.
- BRENER, B. P., CRUZ, M. A., MACEDO, M. S. S. & THOMPSON, R. L. 2024a A highly accurate strategy for data-driven turbulence modeling. *Comput. Appl. Math.* **43**, 59.
- BRENER, B. P., CRUZ, M. A., THOMPSON, R. L. & ANJOS, R. P. 2021 Conditioning and accurate solutions of RANS equations with data-driven turbulence closures. *J. Fluid Mech.* **915**, 1–27.
- BRENER, B. P., MACEDO, M. S. S., THOMPSON, R. L. & IACCARINO, G. 2024b Autoencoding statistics of wall-bounded turbulent flows. (*in preparation*).
- BRES, G. A., BOSE, S. T., EMORY, M., HAM, F. E., SCHMIDT, O. T., RIGAS, G. & COLONIUS, T. 2018 Large-eddy simulations of co-annular turbulent jet using a Voronoi-based mesh generation framework. *AIAA Pap.* 2018-3302.
- CHIN, C., OOI, A., MARUSIC, I. & BLACKBURN, H. 2010 The influence of pipe length on turbulence statistics computed from direct numerical simulation data. *Phys. Fluids* **22**, 115107.
- CRUZ, M. A., THOMPSON, R. L., SAMPAIO, L. E. B. & BACCHI, R. D. A. 2019 The use of the Reynolds force vector in a physics informed machine learning approach for predictive turbulence modeling. *Comput. Fluids* **192**, 104258.
- FONG, C. 2016 Squircular calculations. ArXiv:1604.02174 [math.GM].
- FUKAMI, K., NAKAMURA, T. & FUKAGATA, K. 2020 Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data. *Phys. Fluids* **32**, 095110.
- GUASTI, M. F. 1992 Analytic geometry of some rectilinear figures. *Int. J. Math. Educ. Sci. Technol.* **23**, 895–901.

- KAANDORP, M. L. A. & DWIGHT, R. P. 2020 Data-driven modelling of the Reynolds stress tensor using random forests with invariance. *Comput. Fluids* **202**, 104497.
- KINGMA, D. P. & BA, J. L. 2015 Adam: a method for stochastic optimization. ArXiv:1412.6980 [cs.LG].
- LEE, K. & CARLBERG, K. T. 2019 Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *J. Comput. Phys.* **404**, 108973.
- LING, J., KURZAWSKI, A. & TEMPLETON, J. 2016 Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech.* **807**, 155–166.
- MACEDO, M. S. S., CRUZ, M. A., BRENER, B. P. & THOMPSON, R. L. 2024 A data-driven turbulence modeling for the Reynolds stress tensor transport equation. *Int. J. Numer. Meth. Fl.* **96**, 1194–1214.
- MOHAN, A. T., TRETIAK, D., CHERTKOV, M. & LIVESCU, D. 2020 Spatio-temporal deep learning models of 3D turbulence with physics informed diagnostics. *J. Turbul.* **21**, 484–524.
- MURATA, T., FUKAMI, K. & FUKAGATA, K. 2020 Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *J. Fluid Mech.* **882**, A13.
- NIKURADSE, J. 1930 Untersuchungen über turbulente Strömungen in nicht kreisförmigen Röhren. *Ing. Arch.* **1**, 306–332.
- PINELLI, A., UHLMANN, M., SEKIMOTO, A. & KAWAHARA, G. 2010 Reynolds number dependence of mean flow structure in square duct turbulence. *J. Fluid Mech.* **644**, 107–122.
- PIROZZOLI, S., MODESTI, D., ORLANDI, P. & GRASSO, F. 2018 Turbulence and secondary motions in square duct flow. *J. Fluid Mech.* **840**, 631–655.
- RACCA, A., DOAN, N. A. K. & MAGRI, L. 2023 Predicting turbulent dynamics with the convolutional autoencoder echo state network. *J. Fluid Mech.* **975**, A2.
- SAETTA, E., TOGNACCINI, R. & IACCARINO, G. 2023 Machine learning to predict aerodynamic stall. *Int. J. Comput. Fluid D.* **36**, 641–654.
- THOMPSON, R. L., SAMPAIO, L. E. B., ALVES, F. A. V. B., THAIS, L. & MOMPEAN, G. 2016 A methodology to evaluate statistical errors in DNS data of plane channel flows. *Comput. Fluids* **130**, 1–7.
- VINUESA, R., NOORANI, A., LOZANO-DURÁN, A., KHOURY, G. K. E., SCHLATTER, P., FISCHER, P. F. & NAGIB, H. M. 2014 Aspect ratio effects in turbulent duct flows studied through direct numerical simulation. *J. Turbul.* **15** (10), 677–706.
- VINUESA, R., PRUS, C., SCHLATTER, P. & NAGIB, H. M. 2016 Convergence of numerical simulations of turbulent wall-bounded flows and mean cross-flow structure of rectangular ducts. *Meccanica* **51**, 3025–3042.
- WANG, J.-X., XIAO, H. & WU, J.-L. 2017 Physics informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. *Phys. Rev. Fluids* **2**, 034603.
- WANG, Y., SOLERA-RICO, A., SANMIGUEL VILA, C. & VINUESA, R. 2024 Towards optimal β -variational autoencoders combined with transformers for reduced-order modelling of turbulent flows. *Int. J. Heat Fluid Fl.* **105**, 109254.
- WHITMORE, M., BOSE, S. T. & MOIN, P. 2024 Slip-wall-modeled large-eddy simulation for prediction of turbulent smooth-body separation. *AIAA Pap.* 2024-2374.
- WU, J.-L., XIAO, H. & PATERSON, E. 2018 Physics-informed machine learning ap-

- proach for augmenting turbulence models: a comprehensive framework. *Phys. Rev. Fluids* **3**, 074602.
- WU, J.-L., XIAO, H., SUN, R. & WANG, Q. 2019 Reynolds-averaged Navier-Stokes equations with explicit data-driven Reynolds stress closure can be ill-conditioned. *J. Fluid Mech.* **869**, 553–586.
- WU, X. & MOIN, P. 2008 A direct numerical simulation study on the mean velocity characteristics in turbulent pipe flow. *J. Fluid Mech.* **608**, 81–112.
- ZHANG, B. 2023 Airfoil-based convolutional autoencoder and long short-term memory neural network for predicting coherent structures evolution around an airfoil. *Comput. Fluids* **258**, 105883.