Forecasting extreme dissipation events in wall turbulence using machine learning

By S. R. Harris and M. J. P. Hack

1. Motivation and objectives

Turbulent flows are characterized by enhanced mixing which may increase both heat transfer from walls and viscous losses. A first step towards the control of turbulence is the prediction of peak events which are thought to maintain the chaotic state. A prerequisite is the development of predictive models that forecast the formation of a peak event. Utilizing two separate machine-learning architectures, we explore pathways of estimating the fields associated with peak events on two channel flow simulations at different flow conditions.

In turbulent wall-bounded flow, such as that in the channel, intense and intermittent peak events are ubiquitous (Jiménez 2018). Early simulations by Kim & Moin (1986) suggested that the bursts coincide with the formation of hairpin structures. The detection and prediction of these peak events may provide a pathway for controlling turbulence. Recently, conditional analysis of the high-intensity dissipation events from simulations showed streak and hairpin structures in this chaotic environment (Hack & Schmidt 2021) in developed turbulence. Additionally, proper orthogonal decomposition (POD) showed certain modes that provide energy and dissipation events to the system (Blonigan et al. 2019) and have a similar coherent structure. Hack & Schmidt (2021) showed that the hairpin structures occur at various wall distances within the buffer and logarithmic layers.

In this work we pursue two distinct approaches for the prediction of peak events based on machine learning. In the first case, a convolutional neural network (CNN) is applied in the estimation of the magnitude of the peak dissipation in channel flow at \( \text{Re}_\tau = 176 \) for a time horizon of two eddy turnover times. The second setting considers an approach inspired by natural language processing (NLP) in connection with conditional proper orthogonal decomposition to estimate the evolution of the flow field of peak events in channel at \( \text{Re}_\tau = 2000 \).

2. DenseNet model at low Reynolds number, \( \text{Re}_\tau = 176 \)

In this section we discuss the flows and the machine-learning models for the low Reynolds number channel flow, \( \text{Re}_\tau = 176 \).

2.1. Flow field

An open-source spectral direct numerical simulation (DNS) software (Gibson et al. 2008) was used to simulate a channel flow at \( \text{Re}_\tau = 176 \). Snapshots of the velocity field in the full domain are stored with a frequency of one eddy turnover time, \( \Delta t = h/U_{\text{lam}} \). This frequency also corresponds to a \( \Delta t^+ = 7.75 \). Here the channel half height is \( h = 1 \) and the centerline velocity for a laminar profile is \( U_{\text{lam}} = 1 \). The averaged streamwise velocity profile for this case is shown in Figure 1 and compared to the expected fully turbulent profile from the law of the wall. Subtracting this mean flow from each time instant, we...
Figure 1. (a) Averaged streamwise velocity in the channel for the current case. (b) The solid line shows the DNS simulation of the average streamwise velocity compared to the law of the wall values expected for a fully turbulent channel shown by the dashed line. The $y^+$ is plotted where positive values indicate distance from the wall up to the middle of the channel. The plane of interest at $y^+ = 64$ is also shown as a circle marker for reference.

Figure 2. 2D flow diagram where the wall of the channel is located at the bottom of the flow field and the dotted line indicates the center of the channel. The dashed box identifies an example location of the peak event with the associated origin.

obtain the fluctuation velocity field. Within this work, we focus on a single wall-normal plane at $y^+ = 64$. The definition of the dissipation is

$$D(t, x, y, z) = \nu \frac{\partial u'_i}{\partial x_j} \frac{\partial u'_i}{\partial x_j}, \quad (2.1)$$

where repeated indices imply summation and $u'_i$ indicates the fluctuation velocity in each of the spatial axes, $x_j$, and $\nu = 1/Re$, where the Reynolds number is defined as $Re = U_{\text{lam}} h / \nu$. The local maxima of the dissipation magnitude, $D$, in space and time were found for the specified wall-normal plane. Following the notation by Hack & Schmidt (2021), we also denote the origin of a new axis around these local events. As a consequence, the origin location $(\tilde{t}, \tilde{x}, \tilde{y}, \tilde{z}) = (0, 0, 0, 0)$ denotes the center of the peak event with the dimensions aligning in the same direction as the simulation axes $(t, x, y, z)$. The extent of the volume surrounding the peak event is $(\tilde{L}_x, \tilde{L}_y, \tilde{L}_z) = (1.96, 0.76, 0.98)$. The volume is centered around the domain on the $x, z$ axes, but it contains the complete volume from the wall to a distance of $y^+ = 133$ in the wall-normal direction. This is depicted in Figure 2 for two dimensions. This volume was chosen as it captures the main local flow features of interest.

While the present work considers a single plane, $y^+ = 64$, the self-similarity of the
Forecasting extreme events in wall turbulence using machine learning

Figure 3. (a) Dissipation quantity shown at $y^+ = 64$ distance from the wall. A peak event is shown and indicated by the arrow. This magnitude is the output of the model. The input volume to the model is also indicated by the box at $\tilde{t} = -2\Delta t$ and upstream of the event where $\Delta t$ indicates one eddy turnover time. (b) Distribution of the peak dissipation events in the DNS simulation for $y^+ = 64$.

peak events within the buffer and logarithmic documented by Hack & Schmidt (2021) suggests the suitability of the approach for a range of wall distances.

2.2. DenseNet Model

We now turn our focus to describing a model that predicts the anticipated dissipation magnitude, $D$. Inputs to this model contain the 3D volume of the fluctuation velocity in all three directions at a point two eddy turnover times and upstream from where the peak dissipation events will occur. The output of the model will be the single scalar magnitude of the peak dissipation event that will take place at two eddy turnover times as the center of the domain advects downstream. Figure 3(a) shows the dissipation for a series of time snapshots leading up to the peak event at the wall-normal plane, $y^+ = 64$, of interest. The 3D rectangular volume input location is shown relative to where the peak dissipation events will occur. The origin of the input volume is then shifted to $\tilde{t} = -2\Delta t$ and $\tilde{x} = -1.08$ which corresponds to the 80% of the local mean convective velocity similar to what is done by Hack & Schmidt (2021). We also show in Figure 3(b) the distribution of all the peak events in the simulation study. Of all these events that take place, we train the model to predict on the top 60% largest-magnitude events.

The model tested here is a convolutional neural network (CNN) model that is an extension of the 2D DenseNet architecture initially presented in Huang et al. (2017) to the 3D volume discussed previously. The topology of the 3D DenseNet is shown in Figure 4. This specific architecture was chosen for its ability to feed previous layer information to deeper layers in the network. This residual connections alleviate the vanishing gradient...
Figure 4. (a) A modified Dense Layer that contains layers of LeakyReLU and 3D convolutions of kernel size 1, stride 1, and padding 1 for the first convolution and kernel size 3, stride 1, and padding 1 for the second convolution. (b) A transitional layer connecting the 2 Dense Blocks, consisting of an instance normalization followed by a ReLU activation and 3D convolution of kernel size 1 and stride 1. This was followed by an average pool in 3D of size 2 stride 2 to further reduce to spatial dimensional size. (c) The 3D DenseNet architecture is used in this work and shares significant resemblance to the DenseNet architecture described by Huang et al. (2017). The circles in the Dense Block portion represents a Dense Layer shown in (a). Each Dense Layer outputs 16 channels and each of those are concatenated to the inputs as depicted with the arrows.

The problem of deep architectures allowing for more complex and trainable models. The parameters of the model were trained using a loss function of the mean squared error between the true peak dissipation magnitude and the model

\[
\frac{1}{N} \sum_{n=1}^{N} (\hat{D}_n - D_n)^2,
\]

where \(\hat{D}\) in this equation indicates the model’s predicted dissipation magnitude, \(D\) indicates the true dissipation magnitude from DNS, and \(N\) indicates the number of samples in each training batch. Training of this model used the stochastic optimization method known as Adam (Kingma & Ba 2015), with the weight decay and parameters suggested by Loshchilov & Hutter (2019).

2.3. Results and discussion

The model was trained for several epochs until the loss magnitude was constant for several epochs. Evaluation of the model was based on the coefficient of determination, or \(R^2\), regression score and is presented for the training, validation, and test data shown in Table
Table 1. Model coefficient of determination score trained and tested on the input volume located 2 eddy turnover times upstream of the peak dissipation event.

<table>
<thead>
<tr>
<th>Data set</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>0.95</td>
</tr>
<tr>
<td>Validation</td>
<td>0.45</td>
</tr>
<tr>
<td>Test</td>
<td>0.47</td>
</tr>
</tbody>
</table>

1. An $R^2$ score of 0.95 indicates that 95% of the variation in $D$ can be predicted by the model. An $R^2$ score of 1 means a perfect model, and $R^2$ of zero means a model that is as good as guessing the average value of the outputs without regard to variable inputs. The comparatively large gap in performance between the training and the validation/test sets suggests a certain amount of overfitting occurs in the model. Further refinement of this model to reduce the bias could lead to an improved model. Regularization in the model or augmentation of the training data could be a potential route to increase performance by making the model more generalizable.

An analysis using a saliency map (Simonyan et al. 2014) was also conducted to identify the part of the flow that the model was using to make the predictions. This helps to identify prominent features of the flow field that are important in the prediction. We can observe in Figure 5 that the high-saliency magnitude corresponds to the low-speed streak feature in one of the samples. These flow features are of central importance in making the peak dissipation events distinguishable.

3. Transformer model at high Reynolds number, $Re_T = 2000$

3.1. Flow field

We now turn to the high Reynolds number, $Re_T = 2000$, channel flow case. The data is taken from the DNS by Hoyas & Jiménez (2006) and Lozano-Durán & Jiménez (2014). Similar to the low Reynolds number procedure in identifying peak events, Hack & Schmidt (2021) identified the top 0.1% most intense local maxima, and the reader is referred to that work for further information.

3.2. Transformer model

The model developed here is inspired by architectures commonly used in the field of natural language processing (NLP). Transformers (Vaswani et al. 2017) have attracted significant attention in recent years in applications such as word prediction and machine translation. In the present context, we seek to make predictions of the flow field in a limited domain within the turbulent channel. In NLP applications, the transformer operates on embedded word vectors which represent individual words using a vector of a certain dimension. In our application, we use the modes computed in conditional proper orthogonal decomposition for the embedding of the turbulent flow field.

The flow fields are characterized by the velocity fluctuations in the three Cartesian dimensions as $q = [u', v', w']^T$. The set of input flow fields consists of the 2269 peak events identified in Hack & Schmidt (2021) at a wall location $y^+ = 70$ and an average $\Delta t^+ = 22.23$, which are supplemented by a number of 2560 randomly sampled flow fields at the same wall distance. Taking the instantaneous snapshots as new realizations of the
Figure 5. (a) Streamwise fluctuation velocity at 2 eddy turnover times prior to the peak event at $\tilde{x} = 0$ with the dashed isocontour at $u' = -0.2$ for reference. (b) The matching saliency map to (a). (c) 3D perturbation field of (a) with the same color scale and isocontour. (d) 3D saliency map matching (b). Note here that the wall of the channel is located at the bottom, $y = -1$, of the figures and the flow is traveling in the streamwise, $x$, direction.

Following a similar procedure as outlined by Hack & Schmidt (2021), we compute the singular value decomposition of $Q$ as

$$FQ = U\Sigma V^T,$$

(3.2)

with $F$ being the Cholesky factor of the numerical quadrature weights, $M$, resulting from the discrete approximation. Here, $U$ recovers $n$ eigenvectors of $QQ^TM$ and $\Sigma = \text{diag} [\sigma_1, \sigma_2, ..., \sigma_n]$ are the square roots of the non-zero eigenvalues, $\lambda_i$. This procedure differs from the work by Hack & Schmidt (2021) in avoiding the split into symmetric and anti-symmetric flow fields. The leading eigenfunction again shows the hairpin vortex flow field, $q_i$, we obtain our full velocity matrix

$$Q = [q_1, q_2, ..., q_n].$$

(3.1)
as observed in the varicose modes by Hack & Schmidt (2021), but this is not presented here.

Within the basis described by the POD modes, we can approximate any instantaneous flow field

\[ \tilde{\mathbf{q}} = \mathbf{U}^T \tilde{\mathbf{c}}. \]  

(3.3)

Truncation of \( \mathbf{U} \) to retain the singular modes corresponding to the 2000 largest singular values, this relation becomes an approximation

\[ \tilde{\mathbf{q}} \approx \mathbf{U}^T \tilde{\mathbf{c}}, \]  

(3.4)

which serves as the embedding for our transformer

\[ \mathbf{c} = \mathbf{Uq}, \]  

(3.5)

where \( \mathbf{c} \) is the embedded vector of size \( N_e = 2000 \) for each flow field. The POD-based embedding optimally captures the variance of the underlying data, and distinguishes itself from the embeddings typically used in NLP contexts by the orthogonality of the resulting vectors. Figure 6 shows the reconstruction error as a function of truncation modes, \( N_e \), and marginal error reduction is found for using \( N_e > 2000 \). Similar to the BERT model (Devlin et al. 2019), the present transformer topology is based on the encoder part only. The positional encoding follows the description by Vaswani et al. (2017). The full model is described in Figure 7, where we have used \( n = 2 \) sequences prior to the peak event and \( N_x = 1 \) transformer layers. The hidden dimensions in the feedforward models used a vector size of 2048. A jitter to the input-embedded vectors of 1% and a dropout value of 0.2 was used to allow the model to be more general. Only minor improvement to the model was gained by adding the jitter and dropout rate. Further improvements could include architecture changes such as using decoders instead of encoders only and increasing depth.

The model was trained using a conditioned mean squared error (MSE) where we weight each of the embedded vector outputs by the magnitude of the eigenvalue \( \lambda \). This MSE
allows the model to capture the highest-magnitude eigenfunctions and helps to capture the large structures in the flow. The loss for each output vector is thus defined as

$$ L = \frac{1}{N_e} \sum_{i=1}^{N_e} \lambda_i (c_i - \hat{c}_i)^2, $$

(3.6)

where the subscript $i$ indicates the component of the associated vector.

The inputs to the model contain a sequence of two fluctuation fields that are upstream where the peak event is advected by 80% of the local mean velocity as indicated by Hack & Schmidt (2021) at times $t = -2\Delta t$ and $t = -\Delta t$. These two fields are mapped using the POD embedding to obtain the sequence of embedded vectors $c_1$ and $c_2$. The output of the model is the predicted embedded vector, $\hat{c}$, at the time of the extreme event, $t = 0$, which can be mapped using the embedding to provide the predicted fluctuation field $\hat{q}$ in the volume.

### 3.3. Results and discussion

Training this model on the dataset provided a predictor for the fluctuation field by providing the prior time history of any part of the flow. The MSE loss is shown for the 200 epochs that were computed and the loss is shown for the training and validation datasets in Figure 8. We used the model that obtained the lowest validation MSE loss value, which occurred around epoch number 70. Further training epochs lead to an appreciable overfitting of the model, and further improvements are needed to the architecture and data augmentation to make it more general.

This output of the trained model is compared visually in Figure 9 to what is expected, $q$. The truncation of the eigenmodes to $N_e$ results in a difference in the reconstruction, $\tilde{q}$, as it is now an approximation for the exact fluctuation volume. The model uses the two
Forecasting extreme events in wall turbulence using machine learning

Figure 8. MSE loss as a function of the 200 epochs. Results for the training and validation datasets are indicated.

Figure 9. The streamwise fluctuation field at the point of the peak event. The left column (q) shows the DNS results, the middle column is the reconstructed field using a truncated series of eigenmodes (\(\tilde{q}\)), and the right column shows the predicted result from the trained model (\(\hat{q}\)). The closest wall is near the bottom of the figure.

prior time snapshots and predicts the embedded vector that describes the fluctuation field, \(q\). The results show good agreement for the large-scale features of the fluctuation field.

The peak event presented in the top row of Figure 9 is shown in 3D in Figure 10 with similar colormaps and isocontours for a part of the volume. The results confirms the capturing of the main flow features by the modeling approach.

4. Conclusions

Two machine learning approaches for the forecasting of flow fields in wall turbulence were explored. The first approach was based on a 3D CNN model, whose topology bears some resemblance to the DenseNet architecture, showed significant promise for predicting the magnitude of the peak dissipation event. Slight overfitting of the model was observed, motivating the exploration of improvements based on regularization techniques.
map analysis showed that low speed streaks near the wall are among the significant features used by the model to predict the dissipation magnitude.

The second model was based on a transformer as commonly applied in natural language processing applications. The model used an embedding in terms of POD modes to predict the fluctuation flow field at the time of the peak event. Similar to the CNN model, the training loss shows a certain degree of overfitting that may need to be adjusted in future
work. Preliminary results demonstrate the ability of the model to predict the large-scale features of the turbulent fluctuation field.

Acknowledgments

This work was supported by Sandia National Laboratories. The authors acknowledge use of computational resources from the Certainty cluster awarded by the National Science Foundation to CTR.

REFERENCES


Loshchilov, I. & Hutter, F. 2019 Decoupled weight decay regularization. In ICLR.

