# A weighted sparse-input neural network technique applied to identify important features for vortex-induced vibration

**Leixin Ma, Themistocles L. Resvanis, J. Kim Vandiver**

Department of Mechanical Engineering, Massachusetts Institute of Technology
leixinma@mit.edu

## Abstract

Flow-induced vibration depends on a large number of parameters or features. On the one hand, the number of candidate physical features may be too big to construct an interpretable and transferrable model. On the other hand, failure to account for key dependence among features may oversimplify the model. Feature selection is found to be able to reduce the dimension of the physical problem by identifying the most important features for a certain prediction task. In this paper, a weighted sparse-input neural network (WSPINN) is proposed, where the prior physical knowledge is leveraged to constrain the neural network optimization. The effectiveness of this approach is evaluated when applied to the vortex-induced vibration of a long flexible cylinder with Reynolds number from $10^4$ to $10^5$. The important physical features affecting the flexible cylinders' crossflow vibration amplitude are identified.

## Introduction

Vortex-induced vibration (VIV) is a multi-physics problem associated with a number of features (or variables) that characterize either the structure or the flow individually or their interaction. As flow passes around a cylinder, the wake becomes unstable. The periodically shed vortices induce unsteady forces on the cylinder which lead to VIV. Moreover, the VIV of long cylinders in ocean currents may vary from single mode dominated, narrow-band random vibration to multi-mode response, characterized by broadband random vibration. Different current profiles may cause structural vibration with standing waves or travelling wave patterns (Bourguet et.al. 2011; Vandiver et.al. 2018). The complexity of the nonlinear fluid-structure interaction process, especially for the VIV of long, flexible cylinders in high Reynolds numbers fluid flows, precludes exact analytical solutions and CFD simulations are not yet up to the task.

To identify the key mechanisms and the governing dimensionless parameters behind the complicated fluid-structure interaction process, extensive investigations have been made through structural response measurement, flow visualization and various force modeling techniques (Sarpkaya 2004). VIV research over the past decades has revealed that Strouhal number, Reynolds number, mass ratio, damping parameter etc. are all relevant VIV features (Vandiver 1993, Govardhan and Williamson 2006, Vandiver et.al. 2018). However, if one is only interested in predicting a certain quantity of interest, such as cylinder's vibration amplitude in the crossflow direction, some of these candidate features may be redundant or unimportant.

Feature selection algorithms are intended to extract the most important features out of the full set of candidate features with the goal of keeping prediction accuracy at a desirable level, but with a reduced set of features. A preanalysis of a features' importance can be conducted by examining the statistical correlations among the features. However, the statistical analysis often fails to consider the complicated interactions among the physical input parameters (features). To solve this problem, the importance of each feature subset can be assessed according to their prediction accuracy using a learning machine, such as deep neural network (DNN). Several learning machine-based feature selection approaches have already been developed to iteratively search the optimal feature subset that gives similar prediction accuracy as the full feature set, but they can be computationally expensive especially when the number of input features becomes very large (Guyon 2003).

To efficiently identify important features in a learning machine, several regularization techniques are introduced to the machine learning process. Rudy et. al. (2017) developed a sequential threshold ridge regression, which helped discover governing partial differential equations of a system from measured time series. Inspired by the effectiveness of the group lasso regularization in linear regression, Feng (2017) and Scardapane (2017) developed a sparse-

input neural network by imposing group lasso regularization on the weight groups connecting each input neuron. The effectiveness of the approach was demonstrated through theoretical derivations and empirical evidence.

However, for physical problems, some of the system's properties may be known in advance or can be obtained from the governing physical laws and dimensional analysis (Sonin, 2001). Studies have shown that incorporating the prior physical knowledge can help build more interpretable machine learning models (Ye et. al 2018).

In this paper, the sparse-input neural network proposed by Feng (2017) and Scardapane (2017) is modified to efficiently identify the important features on top of prior physical information. Comparison with searching all combinations of additional features shows its effectiveness in building compact predictive models, while maintaining prior physical information. The method was applied to the VIV response amplitude prediction problem at dominant vibration frequencies. On top of the Reynolds number and damping parameter, the in-line-cross-flow coupling and modal participation are found to be important global VIV features.

## Weighted sparse-input neural network (WSPINN) incorporating prior physical knowledge

We consider a fully connected DNN with $P$ input features $x \in R^P$ in the input layer and $M$ neurons in the first hidden layer that predict a certain target output $y \in R^1$. The weight connecting the $p^{th}$ input feature and $m^{th}$ neuron in the first hidden layer is denoted as $w_{pm}$. Figure 1 shows an example of the DNN with $P$=3 and $M$=4. The sparse-input neural network (Feng 2017; Scardapane 2017) aims at accomplishing two tasks simultaneously: On the one hand, it minimizes $L(y, \hat{y})$, which is the prediction error (or loss) between the predicted $\hat{y}$ and the measured $y$. Meanwhile, it tries to constrain the number of input features to the DNN to be no greater than $k$. To implement this constraint, we need to group the weights outgoing from the same input feature together, and then limit the number of non-zero weight groups to be no larger than $k$. Hence, the mathematical expression for the optimization objective can be expressed as,

$$\min_w L(y, \hat{y})$$

subject to

$$\|w\|_0 = \left|\left\{p : W_p \neq 0\right\}\right| \leq k = \left|\left\{p : \sqrt{\sum_{m=1}^{M}(w_{pm})^2} \neq 0\right\}\right| \leq k \qquad (1)$$

Where $\|w\|_0$ is the $l_0$ norm of weight vector $w$. $|\ldots|$ is the cardinality of the weight groups; $p$ and $m$ are the index for the input feature and the neuron in the first hidden layer, respectively. The magnitude of weight group for feature $p$ is measured by $W_p$. Since the $l_0$ norm is non-convex and non-differentiable, $l_1$ norm, which calculates the sum of absolute values of the vectors, is often used as a convex proxy (Tibshirani, 1996). It can be shown geometrically that $l_1$ norm is the closest convex approximation for $l_0$ norm (Rosasco, 2010). Following the convex approximation, we obtain Equation (1),

$$\min_w \left\{ \frac{1}{N} \sum_{n=1}^{N} L\left(y^{(n)}, \hat{y}^{(n)}\right) + \lambda \sum_{p \in P} \sqrt{\sum_{m=1}^{M}(w_{pm})^2} \right\} \qquad (2)$$

The second term in Equation (2) introduces bias term for prediction. The hyperparameter $\lambda$ is known as the group lasso penalty, which adjusts the sparsity of the input features versus the prediction accuracy. When $\lambda$ grows, the neural network will try to minimize the sum of the weight groups, and therefore more weight groups are likely to shrink to near 0. The input features with nonzero weight groups are the remaining features that contribute to the prediction. In this way, the model can be built out of fewer input features, but the prediction accuracy may decrease due to the loss of information.
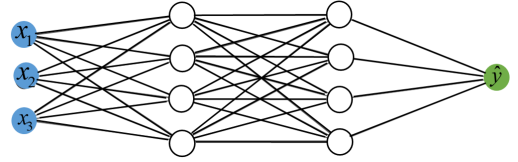


Figure 1: A DNN example with 3 inputs and 4 neurons in the first hidden layer

However, for many physical problems, some of the features are known to be important in advance, which are termed as prior knowledge. In this case, the objective is to select a small number of additional features that will complement the input features that are considered prior knowledge and lead to predictions of acceptable accuracy. Since the conventional sparse-input neural network cannot tell the difference between prior knowledge and additional features, the optimization objective in Equation (2) needs to be modified as follows,

$$\min_w \left\{ L(y, \hat{y}) + s_p \lambda \sum_{p \in P_p} \sqrt{\sum_{m=1}^{M}(w_{pm})^2} + s_a \lambda \sum_{p \in P_a} \sqrt{\sum_{m=1}^{M}(w_{pm})^2} \right\} \qquad (3)$$

Where $P_p$ denotes the feature set representing the prior knowledge, and $P_a$ denotes the set of all the additional features. The parameters $s_p$ and $s_a$ are the weights assigned to the prior knowledge and additional features, respectively. These weights represent the level of confidence on the feature' importance for prediction (Lian, 2018). The conventional sparse-input neural network in Equation (2) is a special case for the weighted formulation in Equation (3),

where $s_a=s_p=1$, which assumes equal confidence for all the features' importance. For the prior knowledge (i.e., feature set known to be important), we'd like to prevent the algorithm from minimizing their weight groups to near 0, hence $s_p/s_a$ should be set close to 0.

# Relevant features for long flexible cylinders subjected to vortex-induced vibrations

## Flexible cylinder VIV modeling

Figure 2 is a sketch of a tensioned elastic cylinder under a linearly sheared current profile $U(z)$ distributed along axis $z$, which causes the cylinders' vibration in both the inline (IL) and crossflow (CF) directions with respect to the incoming current. The vibration of the elastic cylinder can be approximated as a tensioned Euler-Bernoulli beam. The equation of motion in crossflow direction and inline direction can be expressed as,

$$m(z)\frac{\partial^2 y}{\partial t^2}+c_s(z)\frac{\partial y}{\partial t}-P(z,t)\frac{\partial^2 y}{\partial z^2}+EI(z)\frac{\partial^4 y}{\partial z^4}=F_{cf} \quad (4)$$

$$m(z)\frac{\partial^2 x}{\partial t^2}+c_s(z)\frac{\partial x}{\partial t}-P(z,t)\frac{\partial^2 x}{\partial z^2}+EI(z)\frac{\partial^4 x}{\partial z^4}=F_{il} \quad (5)$$

Where $x$ and $y$ are the displacement in inline and crossflow direction. $m(z)$ is the cylinder's mass per unit length, $P(z,t)$ is tension of the vibrating cylinder, $EI(z)$ represents bending stiffness. $c_s$ is the structural damping coefficient per unit length, $F_{cf}$ and $F_{il}$ are the vortex induced forces on the cylinder. The loading transfers energy from fluid to the structure in a well-defined region with length $L_{in}$, which is the "power-in" region. Outside this region, the vortex loading dissipates energy by transferring energy from the structure to the fluid through hydrodynamic damping coefficient $c_h(z)$. The location of the "power-in" region can be identified from structural vibration measurements in experiments or simulation (Rao 2015). Under steady-state, narrow-banded vibration, the total power dissipation in the flexible pipe can be normalized to an equivalent damping coefficient $c_e$ (Vandiver et.al. 2018).
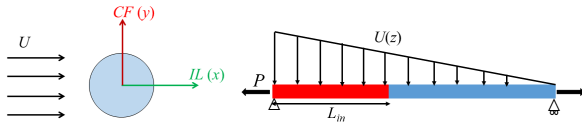


Figure 2: Side and front view of a cylinder under VIV

The VIV loading is the result of nonlinear interaction between vortex shedding and structural vibration via complicated feedback mechanisms that depend on the structural properties, the current profile and the structure's motion

at every instant Hence, the parameterization for VIV force in the "power-in" region may involve,

$$F_{cf}=f(U(z),x(z,t),y(z,t),\rho,\mu,L,L_{in},D,m(z),$$
$$c_{s,cf}(z),c_{h,cf}(z),P(z,t),EI(z)) \quad (6)$$

$$F_{il}=f(U(z),x(z,t),y(z,t),\rho,\mu,L,L_{in},D,m(z),$$
$$c_{s,il}(z),c_{h,il}(z),P(z,t),EI(z)) \quad (7)$$

Where $\rho$, $\mu$, $L$ and $D$ are fluid density and dynamic viscosity, cylinder's length and diameter, respectively.

If the spatiotemporal root-mean-square (rms) amplitude of crossflow vibration $A_{rms,cf}$ in the power-in region is the target output, then from Equations (4)-(7), the predictive model can be expressed as,

$$A_{rms,cf}=f(U(z),x(z,t),y(z,t),\rho,\mu,L,L_{in},D,m(z),$$
$$c_{s,cf}(z),c_{h,cf}(z),P(z,t),EI(z)) \quad (8)$$

It can be observed that Equations (6)-(8) involves spatial-temporal distribution of structural response and system properties, which will be further simplified and represented by some global VIV features.

## Spatial-temporal analysis for typical VIV

VIV measurements from the 2011 Shell experiments on a 38-m-long cylinder (Lie et al 2013) were studied in this investigation.

The measured crossflow displacements in a linearly sheared current are presented in Figure 3. The top figures are the CF response time series at two locations within the "power-in" region, while their corresponding wavelet analysis is shown at the bottom. The vibration is found to be narrow-banded with the dominant frequency $\omega_{cf}$ drifting in time. Given the dominant vibration frequency and structural properties, the corresponding wavenumber $k_{cf}$ can be estimated by the dispersion relationship.
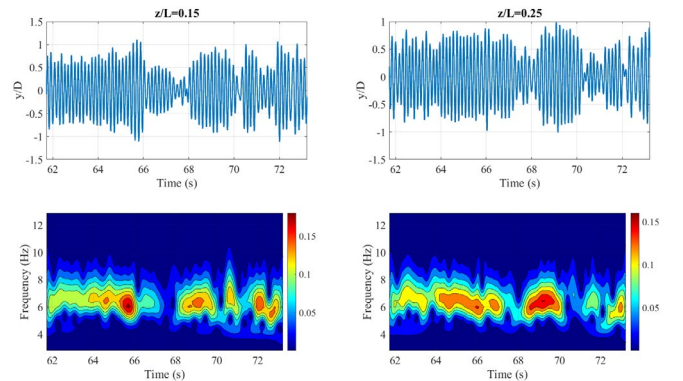


Figure 3: Top: Time series of crossflow displacement at two locations in the power-in region; Bottom: Wavelet analysis on the measured time series. (2011 Shell experiment, $D$=30 mm, linearly sheared flow, $U_{max}$=1.6 m/s)

Meanwhile, Figure 4 shows the corresponding spatial-temporal distribution of crossflow displacement for the same test condition. The response is nonstationary, with a mixture of standing wave and travelling wave components. To better capture the temporal variation of the vibration signal, a moving window analysis is conducted. The vibration signal is windowed into overlapping time frames over each 3 vibration cycles, with 75% overlap.

Complex proper orthogonal decomposition (POD) is conducted on the crossflow displacement in each spatial-temporal window in the "power-in" region to decompose the displacement in each window into several orthogonal complex modes (Feeny 2008). The ratio between the modal energy of the dominant POD mode and the total energy is defined as $\kappa$, which suggests the dominance of the principal mode. Additionally, by comparing the real and imaginary component of the dominant complex mode, the travelling wave index $\alpha$ can be defined, with $\alpha = 1$ for travelling waves, and $\alpha = 0$ for standing waves (Feeny 2008). The middle and bottom of Figure 4 shows the temporal variation of the travelling wave index and the modal dominance factor analyzed in the power-in region, which suggests that the VIV process is single POD mode dominated, but the mode may vary from standing to travelling waves. Analysis from inline vibration also shows similar spatial-temporal distribution.
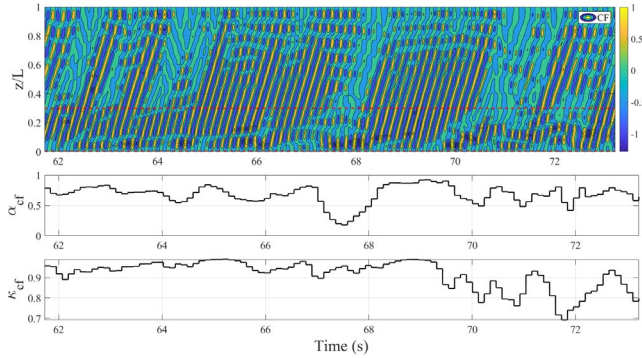


Figure 4: Contour plot of CF vibration around the dominant frequency, its corresponding travelling wave index $\alpha_{cf}$, and mode dominance factor $\kappa_{cf}$ in the estimated "power-in" region (z/L=0~0.3)

## Dimensional analysis for narrow-banded VIV process

For a homogeneous, tensioned cylinder in uniform or linearly sheared current undergoing narrow-banded VIV, Equation (8) can be approximated by the following relevant global quantities,

$$A_{rms,cf} = f(U_{rms}, \Delta U, A_{rms,il}, \omega_{cf}, \omega_{il}, k_{cf}, k_{il}, \alpha_{cf}, \alpha_{il},$$
$$\kappa_{cf}, \kappa_{il}, \rho, \mu, L, L_{in}, D, m, c_s, c_{e,cf}, c_{e,il}, P_0, P, EI) \quad (9)$$

Where $U_{rms}$ and $\Delta U / L$ are the spatial root-mean-square and the shear gradient of the current profile, respectively within the power-in region. $A_{rms,cf}$ and $A_{rms,il}$ are the spatio-temporal rms for the crossflow and inline VIV amplitude in the power-in region. $c_s$ and $c_e$ are respectively, the structural damping coefficient and the equivalent rigid cylinder damping coefficients that will lead to the same power dissipation as discussed by (Vandiver et.al. 2018). $P_0$ and $P$ are the initial tension before VIV and the mean tension during the VIV process, respectively.

Non-dimensionalizing Equation (9) gives,

$$A_{cf}^* = f(\text{Re}, \beta, A_{il}^*, Vr_{cf}, Vr_{il}, Lk_{cf}, \alpha_{cf}, \alpha_{il}, \kappa_{cf}, \kappa_{il},$$
$$L / L_{in}, L / D, m^*\zeta, c_{cf}^*, c_{il}^*, P_0 L^2 / EI, P / (EIk_{cf}^2)) \quad (10)$$

Where $A_{cf}^* = A_{rms,cf} / D$, $A_{il}^* = A_{rms,il} / D$ are the dimensionless crossflow and inline response amplitude, respectively. $\text{Re} = \rho U_{rms} D / \mu$ is Reynolds number; $\beta = (D / U_{rms})(\Delta U / L)$ is known as the shear parameter; $Vr_{cf} = 2\pi U / (\omega_{cf} D)$, $Vr_{il} = 2\pi U / (\omega_{il} D)$ are the crossflow and inline reduced velocities, respectively; $m^*\zeta = 4m\zeta / (\pi \rho D^2)$ is known as the mass damping parameter in the VIV literature, which historically has been thought to be important in controlling rigid cylinder's VIV amplitude. $c_{cf}^* = 2c_{e,cf}\omega_{cf} / \rho U_{rms}^2$ and $c_{il}^* = 2c_{e,il}\omega_{il} / \rho U_{rms}^2$ are the dimensionless forms of the equivalent damping parameter in the crossflow and inline directions, respectively (Vandiver et.al. 2018).

Although Equation (10) suggests that crossflow response prediction in the "power-in" region may require considering the effect of all the 17 dimensionless variables, it is likely that the dimension of the input features can be further reduced due to redundancy or correlation between features or irrelevance to the prediction target. We are interested in finding a smaller and more manageable subset of parameters that are ultimately the most important out of the full set when it comes to determining the CF response amplitude. The motivation behind this is our interest to understand what causes the CF response variability that is observed in the temporal domain. At the very least, we would like to start associating changes to certain parameters with that variability that is often observed but is too complicated to understand.

## Feature selection for flexible cylinder VIV

### Dataset description

The dataset is from a set of experiments conducted by Shell Oil Co. in 2011 at Marintek. The vibration of 38-meter-long cylinders under various current profiles were measured. The test matrix included two pipes with differ-

ent diameters (30-mm and 80-mm) but of the same bending stiffness. The cylinders were tested in uniform and linearly sheared current profiles with the maximum flow speed, $U_{max}$, ranging from 0.5 m/s to 2.5 m/s. This resulted in the Reynolds number Re ranging from $1.0 \times 10^4$ – $20 \times 10^4$. The dataset also included cases where the 80-mm pipe was covered with strakes over 50% of its length. The pipe tests were conducted in uniform flows with $U_{max}$ varying from 0.5 m/s to 1.5 m/s. The strakes dissipated vibration energy and limited the power-in region to $L_{in} = 0.5L$, 50% of the cylinder's length. Detailed descriptions of the experiments can be found in Lie (2013) and Rao (2015).

The structural damping ratio $\zeta$ in the experiment was around 0.5% (Vandiver et.al. 2018). The cross flow reduced velocity $Vr_{cf}$ varies in a narrow range from 6 to 9 and $Vr_{cf} / Vr_{il} \approx 2$.

### Deep neural network setup

The deep neural network was constructed using two hidden layers. Each hidden layer had twenty neurons using a sigmoid activation function. The total number of data points was around 6000. 70% of the experimental data were used as the training data, while the rest was used as the test data. The input variables $x$ were standardized to keep the features at the same scale, while the output variables $y$ were normalized to values between 0 to 1. The mean absolute percentage error (MAPE) was chosen as the loss function between prediction and measurement $L(y, \hat{y})$. The neural network optimization was conducted via FTRL algorithm (McMahan 2013). During neural network training, the batch size was 128 and learning rate was 0.01. The $s_a$ and $s_p$ in Equation (3) were fixed to be 1 and 0.02, respectively. After the optimization, we remove the input features whose magnitude of the weight groups have shrunk to near 0 from prediction model. In this paper, the magnitude of a weight group is considered to be near 0 when it's value is less than 5% of the maximum magnitude among all of the input features.

### Prior physical knowledge for VIV

Experimental studies on small spring-mounted rigid cylinders show that the response amplitude increases with increasing Reynolds number in the range $10^3$ to $10^4$ and decreases as the dimensionless damping increases (Govardhan & Williamson 2006, Vandiver 2012).

Similarly, studies on long flexible cylinders have shown that the Reynolds number and the dimensionless damping continue to play important roles on the VIV response amplitude (Resvanis 2012, Rao 2015) but as discussed earlier, the large number of potentially relevant parameters and the response variability result in scatter in the data.

Because it is known that Reynolds number Re and dimensionless damping parameter $c_{cf}^*$ are important, these

two parameters are designated to be used as prior knowledge. The shear parameter $\beta$ which is ideally suited to differentiating between uniform or sheared flows was the third parameter that was chosen as prior knowledge before starting the feature selections process.

### Feature selection on top of prior physical knowledge

The feature selection procedure was conducted by increasing the hyperparameter $\lambda$ from 0.01 until all the input features except the prior knowledge shrank to 0. Figure 5 demonstrates how varying the value of $\lambda$ determines the number of features chosen by the proposed algorithm. In the figure the retained features are indicated by the presence of a black bar at each $\lambda$ value tested.
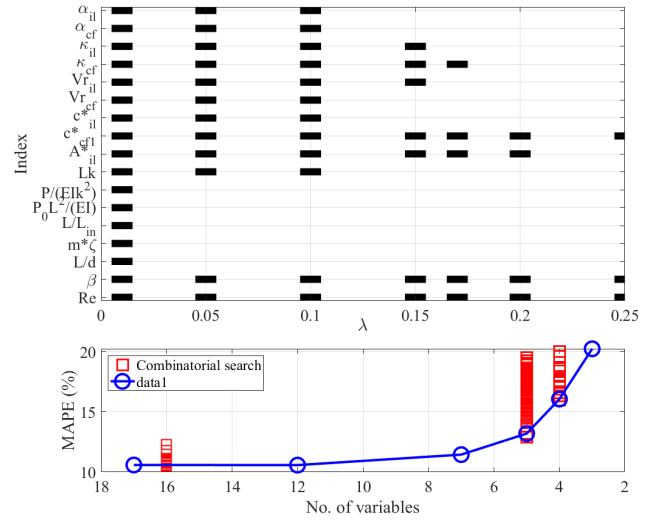


Figure 5: Top: The variation of remaining features with $\lambda$. The black bars represent the features selected by the neural network. Bottom: Comparison of the prediction error (MAPE) between weighted sparse-neural network (WSPINN) prediction and the DNN prediction using combinatorically searched features under a given number of features

The prediction error varies with the retained features in the prediction model, which is presented in the bottom part of Figure 5. At each number of features, a brute force approach that searches all the possible combinations of the additional features is also carried out. The error obtained from the WSPINN is compared with hundreds runs of DNN predictions using combinatorically searched features in addition to the 3 features representing prior knowledge. The comparison suggests that the WSPINN is able to find the feature subsets that gives smallest prediction error among all the feature combinations. Besides, it can be observed that there could be multiple combinations of features that give similar prediction accuracy. For example, both the additional features $A_{il}^*, \kappa_{cf}$ and $A_{il}^*, \kappa_{il}$ gives predic-

tion error around 13%. This suggests the correlations and interactions among some of the VIV features.

After balancing the prediction accuracy with the sparsity of input features, we find that the feature subset containing 5 features: $Re, \beta, c_{cf}^*, A_{il}^*, \kappa_{cf}$ gives 13% MAPE, which is close to 10.6% MAPE using all 17 features.

We have also applied the WSPINN algorithm to other VIV related problems, such as the prediction for rigid cylinder's VIV amplitude and flexible cylinders' VIV amplitude at higher harmonics etc. Because of space limitations we cannot demonstrate this here. Moreover, since this paper only studied the important parameters for VIV sheared and uniform current profiles, the importance of the features may be different for more complicated current profiles.

## Physical insight interpretation

The importance of the identified features for flexible cylinder VIV can be examined by systematically varying the ranges of input features to the constructed neural network models. Figure 6 and Figure 7 show the effect of varying $Re, c_{cf}^*$ while constraining the other variables in the prediction model to characteristic values most often observed in the Shell experiments. The black dots are the experimental measurements within 20% from the referenced values and are included to demonstrate that the prediction model (contours) did in fact have data in that vicinity.

The results demonstrate that increasing Reynolds number tends to increase the spatiotemporal CF RMS amplitude. This Reynolds number effect is obvious in the uniform flow data which typically has small dimensionless damping values ($c_{cf}^*$ <0.3-0.4). While the Reynolds number effect is virtually non-existent when looking at the sheared flow cases with damping parameters ($c_{cf}^*$ >0.4).

Figure 8 shows the effect of varying $A_{il}^*$ and $\kappa_{cf}$ while constraining the other variables. It can be found that the crossflow response tends to increase with inline response. Such a relationship has also been observed in spring-mounted rigid cylinder's VIV experiments (Dahl 2008), where the fluctuating inline force increased with crossflow motion. Finally, the prediction model suggests that as the mode-participation factor increases so does the CF response amplitude. Note that both standing wave and travelling wave response can result in high mode-participation factors and in this situation, the factor primarily characterizes whether all points on the flexible cylinder are responding in a similar manner (spanwise coherence).
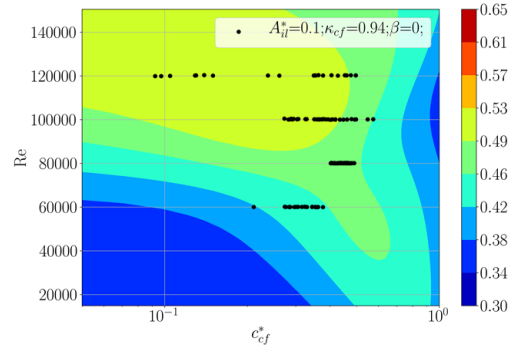


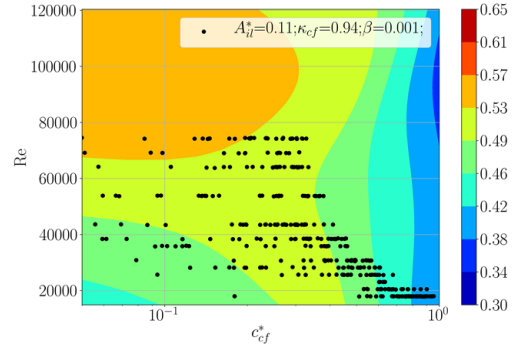Figure 6: Contours of CF RMS amplitude as a function of $c_{cf}^*$ and Re (in uniform flow)



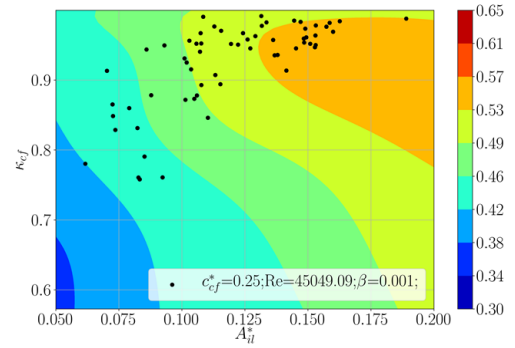Figure 7: Contours of CF RMS amplitude as a function of $c_{cf}^*$ and Re (in sheared flow)



Figure 8: Contours of CF RMS amplitude as a function of $A_{il}^*$ and $\kappa_{cf}$ (in sheared flow)

## Special properties of the approach compared to other machine learning

1. Direct and learning task dependent dimension reduction in the original feature space while retaining the prior information in the model.

The WSPINN is one of the dimension reduction approaches. However, different from widely used PCA or auto-encoders, WSPINN seeks to reduce the dimension directly in the original input feature space. Moreover, through machine learning prediction, WSPINN is able to

identify most important input feature with respect to the target output. The much smaller constraints placed on the prior knowledge also allows the prior knowledge to retain in the prediction model to improve prediction and also identify additional important features.

2. High prediction accuracy due to the universal approximation property of the DNN (Hornik, 1993)

The WSPINN is a feature selection approach embedded in DNN, which is able to predict nonlinear input-output relationships accurately. For instance, for the cross-flow VIV amplitude prediction, the prediction accuracy from the DNN and linear regression given $Re, \beta, c_{cf}^*, A_{il}^*, \kappa_{cf}$ are 13% and 25%, respectively. However, training DNN with WSPINN requires several rounds of iterations to optimize the weights in each layer, hence it was found to be more computationally expensive than most of the other machine learning methods We consider the computational cost acceptable since our intention is not to create a fast predictive tool but rather to use machine learning to reduce the dimensionality of the problem as we try to understand the importance of each of the many governing parameters.

## Conclusion

In this paper, we modify and propose changes to a sparse-input neural network so it can efficiently select additional features which can complement a subset of features known to be important in advance (i.e. prior knowledge). The algorithm was applied to the experimental results from vortex-induced vibration of flexible cylinders. The complicated spatiotemporal response measurements of the continuous system are reduced to an equivalent 2 Degree of Freedom system. The proposed algorithm is then used to investigate the role of Reynolds number, damping parameter, and shear parameter (3 parameters for which we have prior knowledge), as well as 14 other parameters that the dimensional analysis indicated might be important. The algorithm was able to reduce the 14 additional parameters to just 2 additional parameters on top of the prior knowledge. We found that this feature selection technique is much more efficient than a brute force combinatorial search.

## Acknowledgement

## References

Bourguet, R., Karniadakis, G., Triantafyllou, M., 2011. Vortex-induced vibrations of a long flexible cylinder in shear flow. Journal of Fluid Mechanics 677:342-382.

Dahl, J. J. M. 2008. Vortex-induced vibration of a circular cylinder with combined in-line and cross-flow motion (Doctoral dissertation, Massachusetts Institute of Technology).

Feeny, B.F., 2008. A complex orthogonal decomposition for wave motion analysis. Journal of Sound and Vibration, 310(1-2). 77-90.

Feng, J. and Simon, N., 2017. Sparse-input neural networks for high-dimensional nonparametric regression and classification. arXiv preprint. arXiv:1711.07592.

Govardhan, R.N. and Williamson, C.H.K., 2006. Defining the 'modified Griffin plot'in vortex-induced vibration: revealing the effect of Reynolds number using controlled damping. Journal of fluid mechanics. 561, 147-180.

Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of machine learning research, 3(Mar).1157-1182.

Hornik, K., 1993. Some new results on neural network approximation. Neural networks, 6(8), 1069-1072.

Lian, L., Liu, A. and Lau, V.K., 2018. Weighted LASSO for sparse recovery with statistical prior support information. IEEE Transactions on Signal Processing, 66(6). 1607-1618.

Lie, H. et.al., 2013, August. Comprehensive riser VIV model tests in uniform and sheared flow. In ASME 2012 31st International Conference on Ocean, Offshore and Arctic Engineering. 923-930.

McMahan, H.B. et.al., 2013, August. Ad click prediction: a view from the trenches. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.1222-1230. ACM.

Rao, Z., 2015. The flow of power in the vortex-induced vibration of flexible cylinders. Ph. D. Dissertation, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Resvanis, T.L. et.al., 2012, July. Reynolds number effects on the vortex-induced vibration of flexible marine risers. In ASME 2012 31st International Conference on Ocean, Offshore and Arctic Engineering.751-760.

Rosasco, 2010. Statistical Learning Theory and Applications. Lecture notes. Massachusetts Institute of Technology.

Rudy, S.H., Brunton, S.L., Proctor, J.L. and Kutz, J.N., 2017. Data-driven discovery of partial differential equations. Science Advances, 3(4), p.e1602614.

Sarpkaya, T. 2004. A critical review of the intrinsic nature of vortex-induced vibrations. Journal of fluids and structures, 19(4): 389-447.

Scardapane, S., Comminiello, D., Hussain, A. and Uncini, A., 2017. Group sparse regularization for deep neural networks. Neurocomputing, 241: 81-89.

Sonin, A.A., 2001. Dimensional analysis. Technical report, Massachusetts Institute of Technology.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), pp.267-288.

Vandiver, J.K., 1993. Dimensionless parameters important to the prediction of vortex-induced vibration of long, flexible cylinders in ocean currents. Journal of Fluids and Structures, 7(5), 423-455.

Vandiver, J.K., 2012. Damping parameters for flow-induced vibration. Journal of fluids and structures, 35.105-119.

Vandiver, J.K., Ma, L. and Rao, Z., 2018. Revealing the effects of damping on the flow-induced vibration of flexible cylinders. Journal of Sound and Vibration, 433: 29-54.

Ye, T., Wang, X., Davidson, J. and Gupta, A., 2018. Interpretable intuitive physics model. In Proceedings of the European Conference on Computer Vision (ECCV). 87-102.