Estimating a sparse nonlinear dynamical model of the flow around an oscillating cylinder in a fluid flow using SINDy

J. A. Foster [1,2], **J. Decuyper** [1,2], **T. De Troyer** [1,2], **M. C. Runacres** [1,2]
[1] Vrije Universiteit Brussel, Department of Engineering Technology, FLOW
Pleinlaan 2, 1050 Brussels, Belgium
e-mail: **joshua.aiden.foster@vub.be**

[2] Brussels Institute for Thermal-Fluid Systems and Clean Energy
Pleinlaan 2, 1050 Brussels, Belgium

Abstract

The Sparse Identification of Nonlinear Dynamics (SINDy) toolbox can be used to estimate a nonlinear model of dynamical systems. SINDy is a dictionary method that applies sparse regression to a library of candidate functions. The effectiveness of SINDy has been demonstrated in a variety of fields. One such application is the modelling of the wake of a submerged cylinder in a flow. This is regarded as a canonical system for fluid-structure interactions. The SINDy method was found to be successful in modelling the flow around a stationary cylinder. In this work, the technique is applied to the wake of a submerged cylinder undergoing an imposed periodic oscillation. The experiment therefore strongly relates to the case of vortex-induced vibrations (VIV). VIV is challenging to model given that it exhibits much richer nonlinear dynamics than the stationary case. The study is carried out on the vorticity field in the wake of the cylinder. This work demonstrates that SINDy is capable of capturing the observed nonlinear dynamics.

1 Introduction

Of interest is capturing the dynamics of a flow field in the wake of an oscillating cylinder and using this information to obtain a reduced order surrogate model for the system. This is of value as using a reduced order model to make predictions can save valuable time for real-time applications by avoiding the need for repeating costly simulations or experiments.

Thanks to the rapid improvements in computational technology, data-driven modelling techniques have increased in popularity and found their way into many fields of research, among which the study of fluid dynamics. Many fluid-dynamical problems are very complex and high-fidelity predictions can only be achieved by numerically solving the Navier-Stokes equations or by performing experiments. Both methods are time-consuming and costly, which often makes them impractical for applications with limited resources [1].

An example of nonlinear problems that are hard to model are fluid-structure interactions, such as vortex-induced vibrations (VIV). VIV are of particular interest for many applications given that they pose a potential hazard to structural health, especially in regimes of near-resonant frequency locking of the vortex shedding. A canonical system for studying VIV is the flow field in the wake of an oscillating circular cylinder [2]. In such a system, there are several distinct wake regimes that are triggered by the oscillation frequency. Of interest is the lock-in regime, which occurs when the cylinder oscillates near the natural vortex-shedding frequency, also called the Strouhal frequency f_{St}. In this case the vortex-shedding will lock on to the oscillation frequency, resulting in increased loads [1]. The goal of this work is to obtain a data-driven model of such a system that gives insight into the underlying dynamics of the resulting cylinder wake.

The Sparse Identification of Nonlinear Dynamics (SINDy) method is a technique developed by Brunton et al. [3] to identify differential equations that govern the dynamics of nonlinear systems. As such it can be used to provide insight into the nonlinear relationships of different fluid variables. It is a technique that has

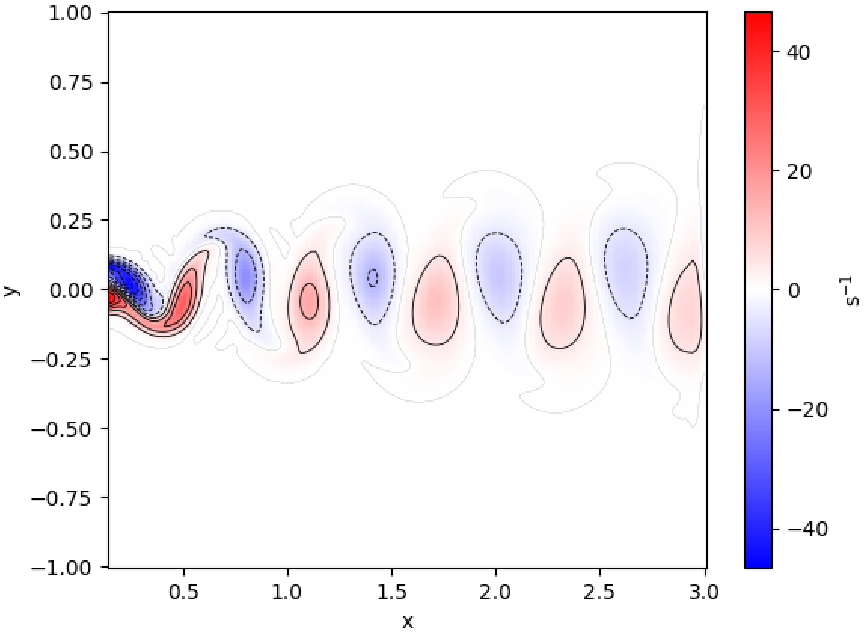


Figure 1: The location of the studied data. From 0.15 to 3 in the x-direction and from -1 to 1 in the y-direction. The central point of the oscillating cylinder is positioned at [0, 0].

been successfully applied to the wake of a stationary cylinder [3]. As the wake of an oscillating cylinder is dependent on more variables, e.g. the flow parameters as well as the cylinder displacement, it will show more complex behaviour than that of a stationary cylinder. As such it is interesting to evaluate the ability of SINDy to model the wake of an oscillating cylinder.

To do this efficiently, an order reduction of the data field is required. SINDy can then be applied to the dynamical evolution of the reduced order state. A Proper Orthogonal Decomposition (POD) order reduction was chosen for its ability to provide insight into the spatial structures of the system. SINDy was then used to identify the dynamics of the corresponding POD coefficients. In this work we worked with vorticity data from a periodic experiment in the lock-in range of the system to enable a meaningful POD-basis.

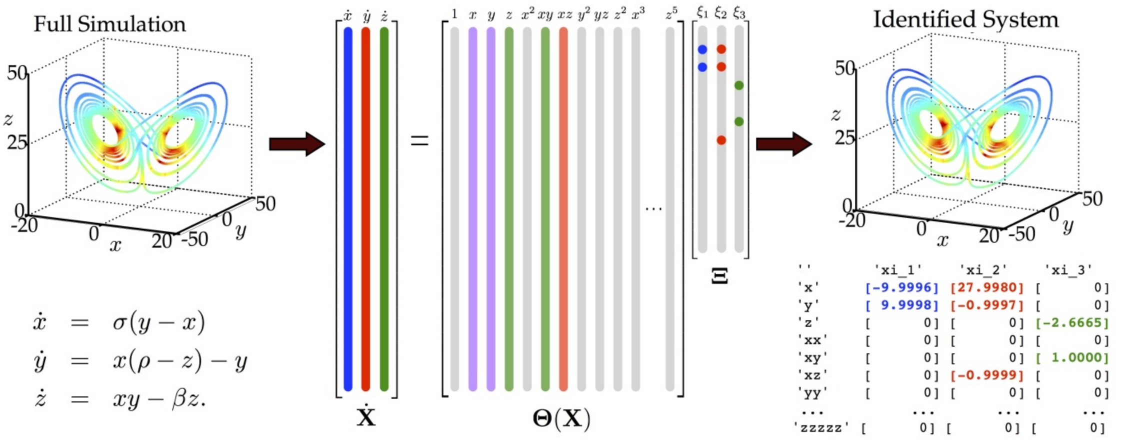
The objective of this work can therefore be summarised as using SINDy to create an ordinary difference equation model that describes the dynamical vorticity field in the wake of an oscillating cylinder in the lock-in regime. The studied system and the data-acquisition are described in Section 2. In Section 3, the SINDy method is discussed. Section 4 demonstrates the resulting model. The results are further discussed in Section 5. Conclusions are formulated in Section 6.

1.1 Notational conventions

Throughout the work, matrices will be denoted by bold faced capital letters, e.g. $\mathbf{A} \in \mathbb{R}^{n \times m}$, vectors are given bold faced lower case letters, e.g. $\boldsymbol{a} \in \mathbb{R}^{n}$, and non-bold lower case letters indicate scalars, e.g. a. Vector functions will be denoted by bold faced italicized letters, e.g. $\boldsymbol{f}()$.

2 Data

In this work, the phenomenon of vortex-induced vibrations is studied from the viewpoint of the vorticity field in the wake of an oscillating cylinder. A 2-dimensional computational fluid dynamics (CFD) simulation of a flow about an oscillating cylinder was used to generate the studied time-series data. The simulation was performed in the open source CFD environment OpenFOAM [4], which implements a finite-volume formulation of the Navier-Stokes equations. A more in-depth discussion on the CFD computations as well


Full Simulation
50
z
25
0
-20
0
x
20
-50
0
y
50
$\dot{x} = \sigma(y - x)$
$\dot{y} = x(\rho - z) - y$
$\dot{z} = xy - \beta z.$
$\dot{x}$ $\dot{y}$ $\dot{z}$
$\dot{\mathbf{X}}$
=
1 x y z x^2 xy xz y^2 yz z^2 x^3 ... z^5
$\mathbf{\Theta}(\mathbf{X})$
ξ_1 ξ_2 ξ_3
Ξ
Identified System

```
''        'xi_1'     'xi_2'     'xi_3'
'x'      [-9.9996] [27.9980] [      0]
'y'      [ 9.9998] [-0.9997] [      0]
'z'      [      0] [      0] [-2.6665]
'xx'     [      0] [      0] [      0]
'xy'     [      0] [      0] [ 1.0000]
'xz'     [      0] [-0.9999] [      0]
'yy'     [      0] [      0] [      0]
...           ...       ...       ...
'zzzzz' [      0] [      0] [      0]
```

Figure 2: An example of SINDy. Sparse regression is performed to fit the library of candidate functions to measurements taken from a system, identifying the model structure used for data generation. **Source:** [6]

as a validation with respect to literature can be found in Decuyper et al. [1].
The flow regime was characterised by a constant Reynolds number of Re = 100, resulting in laminar, mainly 2-dimensional vortex-shedding. During the experiment, the cylinder was forced to oscillate along a trajectory perpendicular to the incoming flow. The oscillation frequency was kept constant at 2.7 Hz, which is 0.9 times the Strouhal frequency of $f_{\mathrm{St}} = 3.0$ Hz. This close proximity to the Strouhal frequency caused the vortex shedding to adjust its natural frequency to the imposed frequency. This is called the lock-in regime [5]. The data therefore contain a regime of periodical vortex shedding at the imposed frequency.

A snapshot of the data that was studied in this work is displayed in Figure 1. With the fluid flowing from left to right in the x-direction, the studied data was selected form 0.15 m to 3.0 m in the x-direction and -1 m to 1 m in the y-direction, relative to the central position of the cylinder. The field was sampled at a frequency of 40 Hz.

3 SINDy

The Sparse Identification of Nonlinear Dynamics (SINDy) is a data-driven modelling technique that was developed by Brunton et al. [3]. Sparsity is promoted considering that the dynamics of most physical systems are governed by only a few relevant terms. In other words, the governing equations of physical dynamics are often sparse when considering a dictionary of nonlinear functions.

With the state of a system at time t denoted by $\boldsymbol{x}(t) \in \mathbb{R}^n$, SINDy is able to identify dynamical models that are defined by the ordinary differential equation (ODE) form

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)), \tag{1}$$

by finding a parsimonious expression for the function $\boldsymbol{f}(\boldsymbol{x}(t))$ that governs the dynamical behaviour of the system states. This is done by first choosing a library of candidate functions of the states, $\mathbf{\Theta}(\mathbf{X})$ and then finding a sparse set of coefficients that correspond to these functions which are stored in the sparse matrix, Ξ, as shown in Figure 2 [3][6].

The method starts from a time series of state measurements, $\mathbf{X} := \begin{bmatrix} \boldsymbol{x}(1) & \boldsymbol{x}(2) & \dots & \boldsymbol{x}(N) \end{bmatrix}^T \in \mathbb{R}^{N \times s}$, with N the number of samples and s the number of states. Both the time derivatives, $\dot{\mathbf{X}}$, and the values of the library functions, $\mathbf{\Theta}(\mathbf{X})$ are calculated for each time step. Then a sparse regression [7][3] is performed to find the coefficients of the most significant terms of $\mathbf{\Theta}(\mathbf{X})$. This results in the sparse coefficient matrix Ξ, where all other coefficients are reduced to zero. Equation 1 then becomes

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)) = \boldsymbol{\Theta}(\boldsymbol{x}(t))\boldsymbol{\Xi}, \tag{2}$$

or, for the whole data set

$$\dot{\mathbf{X}} = \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\Xi}. \tag{3}$$

A more in-depth explanation of this method can be found in [3][6]. Figure 2, demonstrates the method on simulation data extracted from a chaotic Lorenz system. The system is defined by the set of equations shown on the left. In this case a function library of monomials up to the 5th degree is chosen. The sparse coefficient matrix, $\boldsymbol{\Xi} = [\boldsymbol{\xi_1}\ \boldsymbol{\xi_2}\ \boldsymbol{\xi_3}]$, then defines the significance of each monomial when calculating the time derivatives of the states [6]. For example observe from Figure 2 that $\dot{x}$ is constructed using only the first two monomials. Hence only two coefficients are active in ξ_1 (indicated in blue). This corresponds to the governing equation seen on the left. As the example shows, it is possible to identify the ODE structure that was used to generate the data.

In addition to continuous time ODEs, SINDy is also able to generate discrete time ordinary difference equations of the form $\boldsymbol{x}(k+1) = \boldsymbol{f}(\boldsymbol{x}(k))$. In this work discrete-time models are constructed.

3.1 Proper Orthogonal Decomposition

We now know that SINDy can be used to obtain a dynamical model. However, in order to keep the estimation problem tractable, the state variable should remain of moderate size. Note that the two-dimensional time steps of the field data are reshaped into large vectors, $\boldsymbol{x}(t)$, and stored in the data matrix, $\mathbf{X}$. Given that the goal of this work is to model these high-dimensional field data, an order reduction is required.

This can be done using an autoencoder [8], but in this work the choice was made to use a Proper Orthogonal Decomposition (POD) order reduction of the periodic data. This is achieved by projecting the data onto the space spanned by a limited amount of spatial POD modes, $\mathbf{U}$. This is called a truncated POD-basis, $\mathbf{U_r}$. It can be shown that this results in the best possible least squares linear approximation of the data matrix by a lower order matrix [9].

It is important to note that reconstructing the original data matrix from a reduced order approximation will result in a reduction in resolution. This isn't necessarily an unwanted outcome as it enables us to remove noise, but it is important to choose the order high enough to reconstruct the important features.

3.1.1 Order reduction

To reduce the order of the vorticity field 10 spatial POD modes were used. The relative error of the 10 mode reconstruction was 3.87%. The modes are shown in Figure 3 (a). It can be seen that most modes come in pairs with a common feature that is phase shifted by about 90°. The most significant exception to this is the mode that represents the mean field, m. The mode paired with y_1 wasn't considered because it didn't significantly improve the model performance.

The reduced states for each time step, $\boldsymbol{x_r}(t)$ are obtained by projecting the field data, $\boldsymbol{x}(t)$, on the reduced POD-basis:

$$\boldsymbol{x_r}(t) = \mathbf{U_r^{-1}}\boldsymbol{x}(t), \tag{4}$$

with $\mathbf{U_r}$ the reduced set of POD modes and $\boldsymbol{x}(t)$ the snapshots of the data field. The corresponding reduced states as a function of time are shown in Figure 3 (b). Note that the amplitudes of the states decrease from 500 to 20, indicating that the last states correspond to less significant details in the data. Again the phase shift of about 90° is visible for the pairs. As such, alternating between them causes the common features to roughly move along the flow direction as time passes. It is clear that reducing the time-series field data to the evolution of these 10 variables in time is a significant order reduction. The SINDy model will be created with the goal of simulating the evolution of the reduced states through time in a way that provides insight into the underlying dynamics.

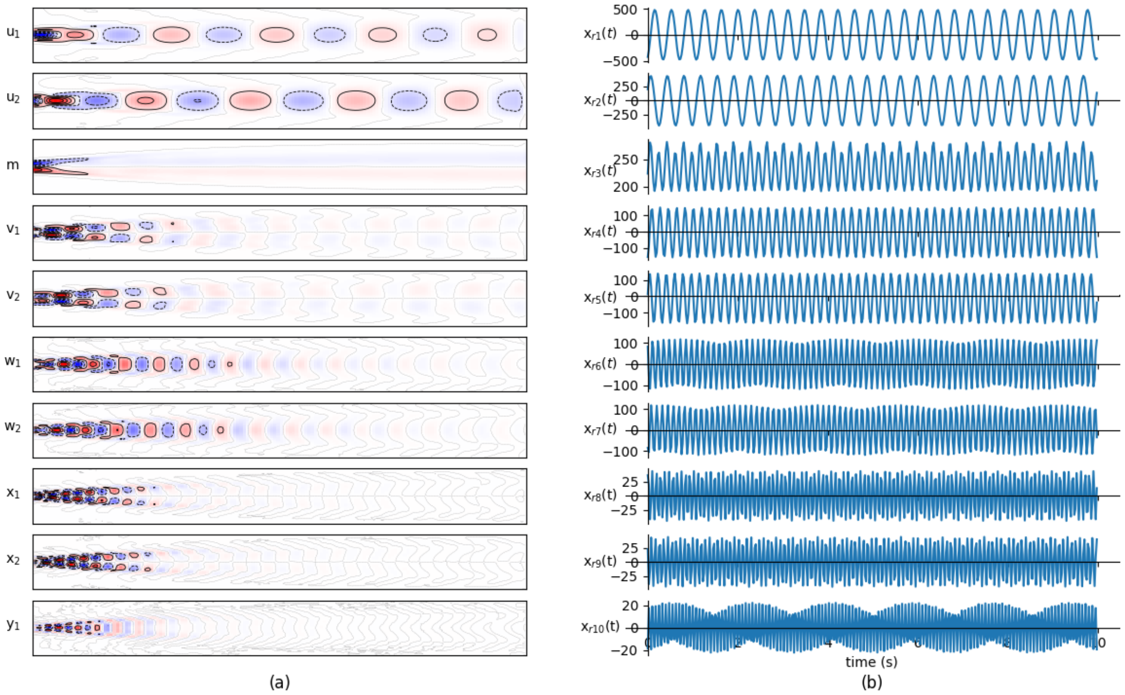


Figure 3: (a) The 10 most significant POD modes of the vorticity field. The modes are cropped in the y-direction for clarity. (b) Shows the projection values of the data onto the corresponding modes.

3.2 Coefficient of performance

As coefficient of performance we will use the relative root mean squared error ($\mathrm{RMSE}_{\mathrm{rel}}$), calculated as the relative Frobenius norm:

$$\mathrm{RMSE}_{\mathrm{rel}} = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|_{\mathrm{F}}}{\|\mathbf{Y} - \overline{\mathbf{Y}}\|_{\mathrm{F}}}, \tag{5}$$

xith $\mathbf{Y}$ the true data matrix, $\widehat{\mathbf{Y}}$ reconstructed data matrix and $\overline{\mathbf{Y}}$ the time average of $\mathbf{Y}$. The overall model performance is given by the $\mathrm{RMSE}_{\mathrm{rel}}$ of the whole time series of $\widehat{\mathbf{X}}$. To evaluate the error evolution through time we can calculate the $\mathrm{RMSE}_{\mathrm{rel}}$ for every individual time step of the data $\hat{\boldsymbol{x}}(t)$. For ease of reading we will simply name this metric the relative error for the remainder of this work.

3.2.1 Error sources

As shown in Figure 4, there are two sources of errors in the reconstruction of the data. The data represented in the original dimensions are indicated in green, while the data converted into the reduced order POD-space are represented in blue. First there will be the resolution loss by using a truncated POD basis. Then the SINDy model, which captures the system dynamics in the reduced order space, will also have an error, the model error. We will define the reconstruction error as the final (combined) error of the reconstruction.

4 Results

The resulting ordinary difference equation model was created using the PySINDy package [10][11] in Python. It captures the dynamics of the wake vorticity while it is represented in the reduced order POD space. The Sparse Relaxed Regularized Regression (SR3) optimiser [12] was used for this application because it consistently resulted in stable models, even when increasing the order from 3 to 10. A polynomial

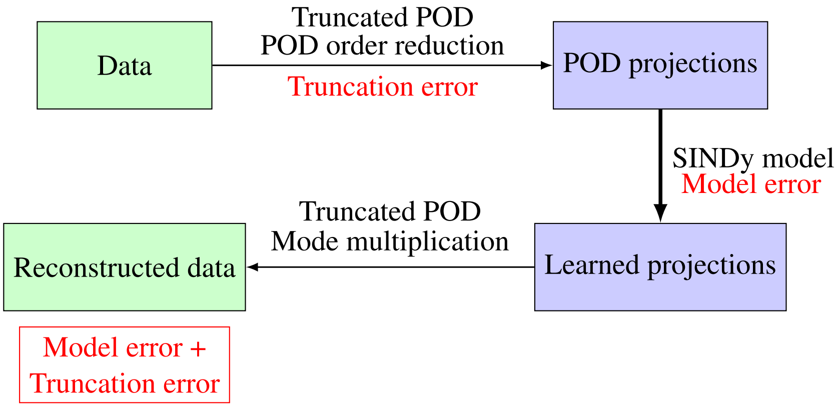


Figure 4: Representation of error sources. Data in the original dimensions are shown in green, data represented in the reduced order POD-space are shown in blue.

dictionary of order 2 was used in combination with a regularisation threshold of 0.035 and with a relaxation value $\mu = 1$. The latter didn't seem to result in any significant changes in the model performance.

4.1 Model structure

The retained monomial terms of the resulting difference equation are shown in the equation below (6). The full equation can be found in Appendix B. For clarity the state variables are named after their corresponding modes, as shown in Figure 3. Studying the equation, it becomes apparent that all second order terms have been removed during the sparse regression, resulting in only first order terms. This is stressed by denoting the functions as f^1 in (6).

$$\begin{aligned}
u_1^{k+1} &= f_1^1\left(u_1^k, u_2^k, m^k, v_1^k, v_2^k, w_1^k, \quad x_1^k, x_2^k, y_1^k\right) \\
u_2^{k+1} &= f_2^1\left(u_1^k, u_2^k, m^k, v_1^k, v_2^k, w_1^k, w_2^k, x_1^k, x_2^k, y_1^k\right) \\
m^{k+1} &= f_3^1\left(u_1^k, u_2^k, m^k, v_1^k, v_2^k, \quad x_1^k, x_2^k\right) \\
v_1^{k+1} &= f_4^1\left(u_1^k, \quad m^k, v_1^k, v_2^k, w_1^k, w_2^k, x_1^k, x_2^k, y_1^k\right) \\
v_2^{k+1} &= f_5^1\left(\quad u_2^k, m^k, v_1^k, v_2^k, w_1^k, w_2^k, x_1^k, x_2^k, y_1^k\right) \\
w_1^{k+1} &= f_6^1\left(u_1^k, \quad m^k, v_1^k, v_2^k, w_1^k, w_2^k, x_1^k, x_2^k, y_1^k\right) \\
w_2^{k+1} &= f_7^1\left(u_1^k, u_2^k, m^k, v_1^k, v_2^k, w_1^k, w_2^k, x_1^k, x_2^k, y_1^k\right) \\
x_1^{k+1} &= f_8^1\left(\quad u_2^k, m^k, v_1^k, v_2^k, w_1^k, w_2^k, x_1^k, x_2^k, y_1^k\right) \\
x_2^{k+1} &= f_9^1\left(\quad m^k, v_1^k, v_2^k, w_1^k, w_2^k, x_1^k, x_2^k, y_1^k\right) \\
y_1^{k+1} &= f_{10}^1\left(\quad m^k, \quad v_2^k, \quad w_2^k, \quad x_2^k, y_1^k\right)
\end{aligned} \tag{6}$$

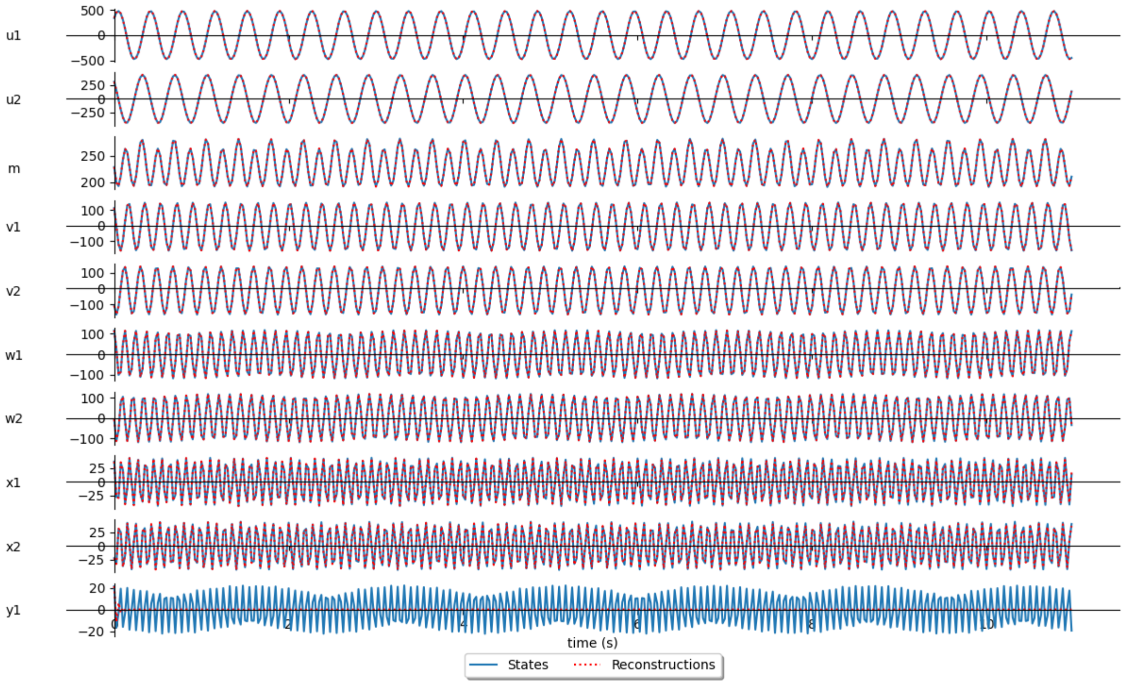


Figure 5: Depiction of the state values through time in blue overlaid by the SINDy model reconstructions in red.

Interesting to note is that, in the full difference equation in Appendix B and the example below, when looking at the dynamics of each mode pair, e.g. u_1 and u_2, the paired modes have a similar effect on their own dynamics. As demonstrated in (7) and (8), respectively $u_1^{k+1} \propto 0.916\ u_1^k$ and $u_2^{k+1} \propto 0.911\ u_2^k$. On the other hand their effects on each other are approximately opposite, $u_1^{k+1} \propto +0.427\ u_2^k$ in (7) while $u_2^{k+1} \propto -0.397\ u_1^k$ in (8):

$$f_1^1 = u_1^{k+1} = 0.916\ u_1^k + 0.427\ u_2^k + \dots \tag{7}$$

$$f_2^1 = u_2^{k+1} = -0.397\ u_1^k + 0.911\ u_2^k + \dots \tag{8}$$

The effects of the other modes on the pair are usually much lower. This becomes more distinct in the functions describing the subsequent pairs.

4.2 Model performance

The evolution of the modes and their reconstructions by the SINDy model are depicted in Figure 5. It is clear that all states except for y_1 are modelled accurately, while y_1 decays rapidly.

The evolution of the relative errors through time is given in Figure 6. It is interesting to note that the truncation error and the model error are approximately in counter-phase, resulting in a beneficial accumulated error when reconstructing the data.

As shown in Table 1, the 10-mode POD-representation resulted in a relative reconstruction error of 0.0387. The error of the SINDy model in capturing the system dynamics in the POD-space was 0.0310. The overall error of reconstructing the vorticity field from the model predictions is 0.0504.

5 Discussion

We have shown that a discrete time SINDy model can accurately capture the dynamical behaviour of the vorticity field in the wake of a periodically oscillating cylinder when projected onto a 10-mode POD-space.

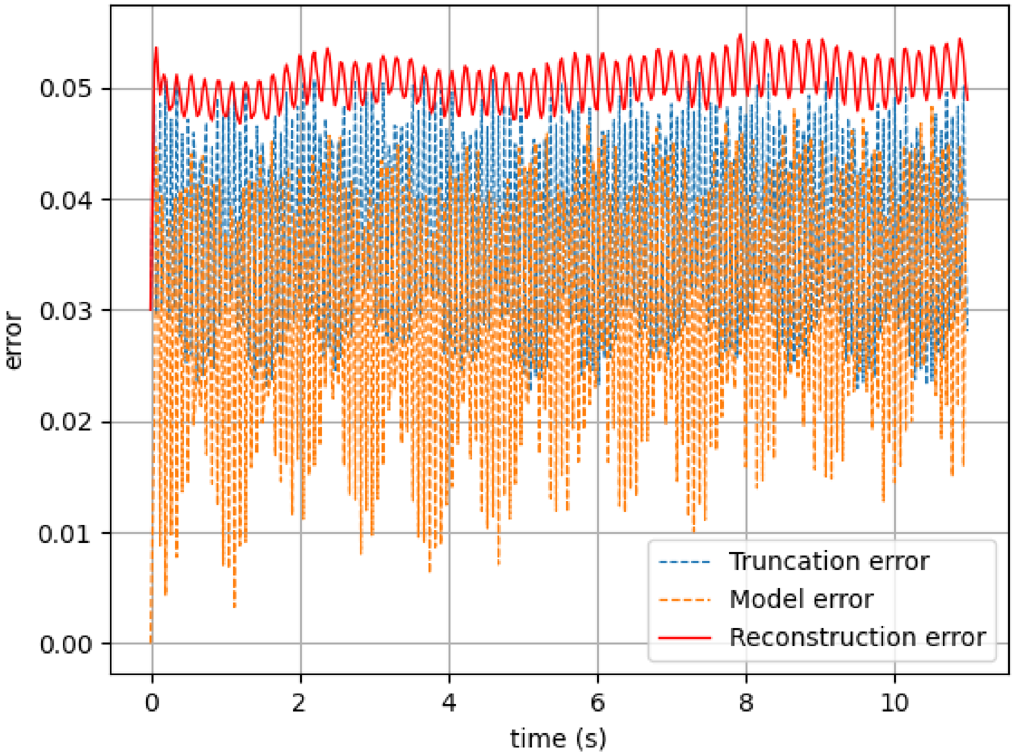


Figure 6: Evolution of the relative error sources through time. The errors caused by the POD truncation and the SINDy model are shown in dashed lines. The final reconstruction error is displayed as a full red line.

Table 1: Relative errors

	Relative Frobenius norm
POD truncation	0.0387
Model error	0.0310
Reconstruction error	0.0504

Firstly it is interesting to note that the model error is in counter-phase with POD truncation error, adding up to the reconstruction error of 5 %. This is in the same order of magnitude as the uncertainty on the CFD data themselves, meaning that lower reconstruction errors would not be meaningful.

Secondly the decay of the state corresponding to y_1 indicates that the priority of modelling this state accurately was low. The high amount of terms needed to model this state in the absence of its paired state in combination with its low significance compared to the other states resulted in the sparse regression diminishing this state.

Additionally, the removal of all second order monomials resulted in a set of only first order difference equations. This indicates that the periodic data that was studied in this work provides a highly linear viewpoint into the overall nonlinear dynamical system when using the POD-basis. As such, a linear model could be constructed that was able to capture the behaviour of this nonlinear system in a limited regime. Further work will be needed to expand to a range of system regimes.

It is also interesting to note that the periodicity of the vortex shedding is clearly reflected in the relationships in the POD-mode pairs. A positive or negative value for one respectively causes its paired mode to grow or decrease while the opposite holds for the influence it feels from its paired mode. At the same time they influence themselves with similar strengths. This results in the paired modes alternately growing and decaying. Combined, this roughly causes their corresponding feature to move in the direction of the flow as time progresses. As such, the model is able to provide a level of insight into the system.

6 Conclusion

In this work the dynamical behaviour of the periodic vorticity field in the wake of an oscillating cylinder was identified by training a SINDy model on the 10-mode POD-space projection of the spatiotemporal field data. The model captured the dynamics with a relative error of 3.1 % and yielding an overall data reconstruction with a relative Frobenius norm of 5 %. The difference equation structure of the model made it possible to draw conclusions about the studied system, such as its linearity and periodicity.

References

[1] J. Decuyper, T. De Troyer, K. Tiels, J. Schoukens, and M. Runacres, "A nonlinear model of vortex-induced forces on an oscillating cylinder in a fluid flow," *Journal of Fluids and Structures*, vol. 96, 7 2020.

[2] J. Foster, J. Decuyper, M. Runacres, and T. De Troyer, "Modelling vortex-induced loads using recurrent neural networks," in *Modeling, Estimation and Control Conference MECC 2021*, ser. IFAC Proceedings Volumes, J. Wang, H. Fathy, Q. Wang, and B. Ren, Eds., vol. 54, no. 20. IFAC - PapersOnLine, 10 2021, pp. 32–37.

[3] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.1517384113

[4] C. J. Greenshields, *OpenFOAM User Guide*, 4th ed., OpenFOAM Foundation Ltd., June 2016. [Online]. Available: http://openfoam.org

[5] S. Kumar, Navrose, and S. Mittal, "Lock-in in forced vibration of a circular cylinder," *Physics of Fluids*, vol. 28, no. 11, p. 113605, 2016. [Online]. Available: https://doi.org/10.1063/1.4967729

[6] S. L. Brunton and N. J. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems and Control.* Cambridge University Press, 2017.

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x

[8] K. Champion, B. Lusch, J. Kutz, and S. Brunton, "Data-driven discovery of coordinates and governing equations," *Proceedings of the National Academy of Sciences*, vol. 116, p. 201906995, 10 2019.

[9] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936. [Online]. Available: https://doi.org/10.1007/BF02288367

[10] B. de Silva, K. Champion, M. Quade, J.-C. Loiseau, J. Kutz, and S. Brunton, "Pysindy: A python package for the sparse identification of nonlinear dynamical systems from data," *Journal of Open Source Software*, vol. 5, no. 49, p. 2104, 2020. [Online]. Available: https://doi.org/10.21105/joss.02104

[11] A. A. Kaptanoglu, B. M. de Silva, U. Fasel, K. Kaheman, A. J. Goldschmidt, J. Callaham, C. B. Delahunt, Z. G. Nicolaou, K. Champion, J.-C. Loiseau, J. N. Kutz, and S. L. Brunton, "Pysindy: A comprehensive python package for robust sparse system identification," *Journal of Open Source Software*, vol. 7, no. 69, p. 3994, 2022. [Online]. Available: https://doi.org/10.21105/joss.03994

[12] P. Zheng, T. Askham, S. L. Brunton, J. N. Kutz, and A. Y. Aravkin, "A unified framework for sparse relaxed regularized regression: Sr3," *IEEE Access*, vol. 7, pp. 1404–1423, 2019.

Appendix

A Nomenclature

The next list describes several acronyms and symbols that will be used within the body of this work. Matrices are denoted by bold faced capital letters, e.g. $\mathbf{A} \in \mathbb{R}^{n \times m}$, vectors are given bold faced lower case letters, e.g. $\boldsymbol{a} \in \mathbb{R}^n$, and non-bold lower case letters indicate scalars, e.g. a. Vector functions will be denoted by bold faced italicized letters, e.g. $\boldsymbol{f}()$.

A.1 Acronyms

CFD	Computational Fluid Dynamics
POD	Proper Orthogonal Decomposition
RMSE_{rel}	Relative Root Mean Squared Error
SINDy	Sparse Identification of Nonlinear Dynamics
SR3	Sparse Relaxed Regularized Regression

A.2 Symbols

t	time (s)
$\mathbf{X}$	measured system state matrix
$\dot{\mathbf{X}}$	time derivative of the system state matrix
$\widehat{\mathbf{X}}$	reconstructed system state matrix
$\overline{\mathbf{X}}$	time average of the measured system state matrix
$\boldsymbol{x}(t)$	measured system state vector at one time step
$\dot{\boldsymbol{x}}(t)$	time derivative of the measured system state vector at one time step
$\mathbf{U}$	matrix containing the POD-modes
$\mathbf{U_r}$	matrix containing the reduced or truncated POD-modes
$\boldsymbol{x}_{\mathbf{r}}(t)$	reduced state vector at one time step
$f()$	function
$\boldsymbol{f}()$	vector function
$\boldsymbol{\Theta}()$	function library matrix
$\boldsymbol{\Xi}$	function coefficient matrix
$\boldsymbol{\xi}$	function coefficient vector for single state
f_{St}	Strouhal frequency (Hz)

B Full model

$$\begin{aligned}
u_1^{k+1} &= -260.936 + 0.916\,u_1^k + 0.427\,u_2^k + 1.119\,m^k - 0.135\,v_1^k - 0.262\,v_2^k - 0.033\,w_1^k + 0.019\,x_1^k \\
&\quad + 0.011\,x_2^k - 0.004\,y_1^k \\
u_2^{k+1} &= 12.147 - 0.379\,u_1^k + 0.911\,u_2^k - 0.052\,m^k + 0.013\,v_1^k + 0.012\,v_2^k - 0.023\,w_1^k - 0.019\,w_2^k \\
&\quad - 0.001\,x_1^k + 0.001\,x_2^k + 0.003\,y_1^k \\
m^{k+1} &= 21.803 - 0.007\,u_1^k - 0.001\,u_2^k + 0.902\,m^k - 0.202\,v_1^k + 0.034\,v_2^k + 0.017\,x_1^k - 0.013\,x_2^k \\
v_1^{k+1} &= 23.472 + 0.001\,u_1^k - 0.073\,m^k + 0.674\,v_1^k + 0.780\,v_2^k + 0.008\,w_1^k + 0.003\,w_2^k - 0.026\,x_1^k \\
&\quad + 0.004\,x_2^k + 0.002\,y_1^k \\
v_2^{k+1} &= -60.423 - 0.003\,u_2^k + 0.222\,m^k - 0.766\,v_1^k + 0.603\,v_2^k - 0.006\,w_1^k - 0.002\,w_2^k - 0.026\,x_1^k \\
&\quad - 0.020\,x_2^k + 0.003\,y_1^k \\
w_1^{k+1} &= -6.426 - 0.001\,u_1^k + 0.026\,m^k - 0.010\,v_1^k - 0.008\,v_2^k + 0.292\,w_1^k + 0.959\,w_2^k + 0.004\,x_1^k \\
&\quad + 0.001\,x_2^k - 0.001\,y_1^k \\
w_2^{k+1} &= 6.160 - 0.001\,u_1^k - 0.002\,u_2^k - 0.027\,m^k + 0.012\,v_1^k + 0.008\,v_2^k - 0.953\,w_1^k + 0.293\,w_2^k \\
&\quad + 0.002\,x_1^k + 0.001\,x_2^k - 0.004\,y_1^k \\
x_1^{k+1} &= -25.873 - 0.002\,u_2^k + 0.110\,m^k - 0.017\,v_1^k - 0.026\,v_2^k + 0.001\,w_1^k - 0.002\,w_2^k - 0.126\,x_1^k \\
&\quad + 0.997\,x_2^k + 0.004\,y_1^k \\
x_2^{k+1} &= -1.335 + 0.006\,m^k + 0.001\,v_1^k - 0.004\,v_2^k - 0.001\,w_1^k + 0.002\,w_2^k - 0.987\,x_1^k - 0.128\,x_2^k \\
&\quad - 0.001\,y_1^k \\
y_1^{k+1} &= 1.537 - 0.007\,m^k + 0.001\,v_2^k - 0.001\,w_2^k - 0.002\,x_2^k - 0.527\,y_1^k
\end{aligned}$$