27th ABCM International Congress of Mechanical Engineering
December 04-08, 2023, Florianópolis-SC, Brazil

COB-2023-1366
MODELING A DYNAMICAL SYSTEM FROM LOW-SAMPLED DATA

Davi Saadi de Almeida Lettieri
Leonardo Santos de Brito Alves
Departamento de Engenharia Mecânica, Universidade Federal Fluminense, Niterói, RJ 24210-240, Brasil.
davilettieri@id.uff.br
lsbalves@id.uff.br

Abstract. *The derivation of differential equation models for experimentally observed dynamics has always been a challenge in any field of science. In general, two approaches are employed to do so. One derives such models from first-principles using a heavy dose of mathematical manipulation. In this physics-driven approach, experimental data is employed to validate the model's accuracy. Another is based on recent developments in measurement capabilities, which can provide a wealth of information that allows optimization methods to reverse engineer these models. Arguably the most prominent method within this data-driven approach is the Sparse Identification of Nonlinear Dynamics, i.e. SINDy. It takes advantage of data wealth and model sparsity to employ linear regression techniques to cheaply estimate the nonlinear differential model. Since its release, many expansions have made SINDy even more robust and flexible. However, some obstacles still exist when trying to use this tool on experimental data. In particular, low-sampled data makes the calculation of the time derivatives required by SINDy too inaccurate for it to provide a valid model. In order to overcome this, a sample-rearranging technique was developed for low-sampled periodic data sets that span many periods. The idea is based on the fact that, even for a low sampling rate, the phase portrait topology contains all the relevant information for a periodic nonlinear regime. This technique accurately reconstructs the temporal dynamics of a single period by accordingly rearranging the available data in all other periods into it using the dynamics characteristic period. In the present paper, the method's efficacy is tested using an artificially low-sampled data set from the Lorenz equations for periodic parametric conditions. SINDy works, but only when used in conjunction with the sample-rearranging.*

Keywords: *Reduced Order Model, Low-Sampled Data, SINDy, Sample-Rearranging.*

1. INTRODUCTION

There is a continuous demand for new approaches that can identify differential equations that correctly describe a phenomenon of study. That is especially true in fluid mechanics since many fluid flows represent a high-dimensional nonlinear behavior. However, it is possible to decompose the very complex observed dynamic into a few dominant coherent structures representing most of the total system's energy (Holmes *et al.*, 1996). One of the most common methodologies for dimensionality reduction is the proper orthogonal decomposition (POD) (Sirovich, 1987). The POD is only one of the many techniques within the Machine Learning (ML) field and can be combined with others to extract even more information from a data set (Brunton *et al.*, 2020), as building a reduced order model (ROM). The works of Bongard and Lipson (2007) and Schmidt and Lipson (2009) must be highlighted as milestones in this field since both successfully described a data-driven methodology that automatically generates a system of differential equations using symbolic regression. While previous works that also used this type of regression were limited to linear systems or manual insertion of nonlinear models, employing Genetic Program (Koza, 1992) was the key to reproducing nonlinear coupled dynamical systems. However, the Sparse Identification of Nonlinear Dynamics (SINDy) is the focus of this work. Recently proposed by Brunton *et al.* (2016), it forces the generation of an interpretable and parsimonious ROM by a cheap sparse symbolic regression, simultaneously avoiding overfitting. The regression takes the form of a minimization problem in which sparsity is enforced by different methods. In the original paper, the authors opted for the sequential threshold least squares (STLSQ), while the LASSO (Tibshirani, 1996) was also advocated. Since then, an additional set of alternative methodologies have been described (Boninsegna *et al.*, 2018; Zhang and Lin, 2018; Zheng *et al.*, 2018; Champion *et al.*, 2020). Other works have contributed to SINDy's improvement through framework modifications, allowing its applications to a wealthier range of nonlinear behaviors. The first extension of SINDy concerns the ability to produce equations with rational functions, the implicit-SINDy (Mangan *et al.*, 2016). After that, the works of Rudy *et al.* (2017) and Schaeffer (2017) show how to discover a system of partial differential equations. Another work focused on the minimization problem generalization: the data-driven Galerkin regression (Loiseau and Brunton, 2018). Recently, the development of the SINDy-PI (parallel and implicit) (Kaheman *et al.*, 2020) synthesized and optimized all the previously described expansions into a single tool.

The range of works that employed SINDy is too large to be cited here. However, properly building the matrix of derivatives is fundamental and one of the biggest challenges when using SINDy. There are different ways to approach this problem. Finite differences would be intuitive, but the noise presented in the data can be over-amplified, making this methodology prohibitive. In the original SINDy paper, the authors advocate for the total-variation regularized derivative. Based on previous experiences in our group we suggest first finding an analytical function for the original data via least-squares regression and then evaluating its derivative (Lettieri *et al.*, 2022). Here, our focus lies on the problem where the data set available was not taken at a high enough rate to capture all the dynamics. This is very common, especially when working with experimentally obtained data. Although the behavior associated with the fundamental frequency appears in the original data, often, the higher harmonics ones do not. However, noticing that the phenomenon of interest is periodic, the 2D phase portrait can present all the information needed. Based on this, we developed the sample-rearranging technique, a tool capable of reorganizing an initial low-sampled data set. After being subjected to the rearranging process, the available low-sampled data points are reallocated into a single period domain, simulating a high-rate sampling measurement. Therefore, it reduces the effort when performing the least squares regression and, at the same time, helps generate the analytical functions that reproduce all the important dynamics and are used to build the matrix of derivatives.

In the past, the Energy and Propulsion Research Laboratory at UCLA (University of California, Los Angeles) has been studying the flame dynamics produced by the coupled combustion of individual droplets for different liquid fuels under different external acoustic excitation (Sevilla-Esparza *et al.*, 2014; Karagozian, 2016; Bennewitz *et al.*, 2018). Recently, the research changed focus to the combustion of laminar gaseous jets emanating from single microjets (Sim *et al.*, 2020). From these studies, one can learn about the flame shape and structure and the transitions present in the combustion process dynamics. These works are based on high-speed time-resolved images obtained from a well-established experimental apparatus (Sevilla-Esparza *et al.*, 2014; Bennewitz *et al.*, 2018; Sim *et al.*, 2020; Vargas *et al.*, 2020; Vargas, 2022 - 2022). The jet Reynolds number Re, the acoustic forcing frequency f_a, and the pressure amplitude perturbation that characterizes the amplitude of acoustic excitation can vary. To better understand the flame dynamics, our goal is to construct a ROM using a Snapshot POD from the available images. Unfortunately, this case belongs to a previously mentioned class of problems in which the experimental gear available, although powerful, is still limited to a relatively low imaging frame rate, providing only a sample unable to capture the higher harmonics dynamics. Therefore, we hope that our method can overcome this issue.

2. METHODOLOGY

2.1 SINDy

Consider the system of ordinary differential equations in the form of

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t)) \ , \tag{1}$$

where $\mathbf{x}(t) = \{x_1(t), x_2(t), \ldots, x_n(t)\}$ and $\mathbf{f} = \{f_1(\mathbf{x}(t)), f_2(\mathbf{x}(t)), \ldots, f_n(\mathbf{x}(t))\}$ are the state vector and function, respectively. To use SINDy, we first need a time-data series organized as a data matrix $\mathbf{X} = [\mathbf{x}(t_1), \mathbf{x}(t_2), \cdots, \mathbf{x}(t_m)]$, where m is the sampling rate over the period $\tau = t_m - t_1$. The next step is building the derivative matrix

$$\dot{\mathbf{X}} = [\dot{\mathbf{x}}(t_1), \dot{\mathbf{x}}(t_2), \cdots, \dot{\mathbf{x}}(t_m)] \ . \tag{2}$$

Then, we propose a library of candidate nonlinear functions $\mathbf{F}(\mathbf{X}) = \begin{bmatrix}\mathbf{1} & \mathbf{X}^1 & \mathbf{X}^2 & \cdots & \mathbf{X}^d\end{bmatrix}$, where the sub matrix $\mathbf{X}^d$ is formed by column vectors representing one of all possible polynomial combinations of d^{th}-degree for all times. To quantify how active each function is for each row of Eq. (1) we define the coefficient matrix $\mathbf{C} = [\dot{\mathbf{c}}_1, \dot{\mathbf{c}}_2, \cdots, \dot{\mathbf{c}}_n]$, where for a equation $k = 1, 2, \ldots, n$ of the system defined by Eq. (1) we would have to solve the problem

$$\dot{\mathbf{X}}_k = \mathbf{F}(\mathbf{X})\mathbf{C}_k \ . \tag{3}$$

Since, usually Eq. (3) is an over-determined system, we must solve the following optimization problem

$$\mathbf{C}_k = \underset{C_k'}{\text{argmin}} \ \|\dot{\mathbf{X}}_k - \mathbf{F}(\mathbf{X})\mathbf{C}_k^{'}\|_2 - \lambda\|\mathbf{C}_k^{'}\|_1 \ , \tag{4}$$

where λ is a regularization parameter used to promote sparsity if we choose to use the LASSO (Tibshirani, 1996) in a sparse regression. However, this work employs the sequential thresholded least-squares (STLSQ) method. The latter is not only advocated in other works (Brunton and Kutz, 2022; Champion *et al.*, 2020) but our experience suggests a better performance to select the correct set of candidate function for the reduced order model. Thus, one must define the lowest magnitude a coefficient v_{crit} is allowed to be so its associated candidate function is not dropped from the next regression iteration. Finally, our system of ordinary differential equations can be written as $\mathbf{x}_k = \mathbf{F}(\mathbf{x}^T)\mathbf{C}_k$, where $\mathbf{F}(\mathbf{x}^T)$ is a vector of symbolic functions.

2.2 Sample-Rearranging

Many times, we focus on employing the SINDy on experimentally obtained data of a particular nonlinear dynamics that exhibits a periodic behavior. Often, however, the gear available can only produce a low-sampled data set and we are not able to work with a time series taken at a high enough rate. Hence, even if the Nyquist-Shannon sampling theorem (Nyquist, 1928; Shannon, 1948) is satisfied for the fundamental frequency of the periodic regime, we can not capture the effect of the higher harmonics when trying to reproduce the observed phenomenon. This makes the construction of the derivative matrix in Eq. 2 particularly difficult since we choose to find the approximated functions that better fit the original time series first and only then build it.

To overcome this issue, we propose the sample-rearranging technique. Figure 1 illustrates the fundamental idea of the tool. Disposing of a low-sampled data set with several points scattered along different periods over the original time domain, we seek to compress them into a single period. Taking advantage of the periodic behavior, we construct the algorithm to preserve the same information that a particular point offered before the arranging process. Thus, we dynamically back the future points off to the first one immediately after we identify the number of periods a point is further in time. Therefore, we simulate a scenario where we generate a data set in which each point is artificially taken at a significantly higher rate. Thus, producing a set of fitted functions that capture the higher harmonics dynamics becomes possible.

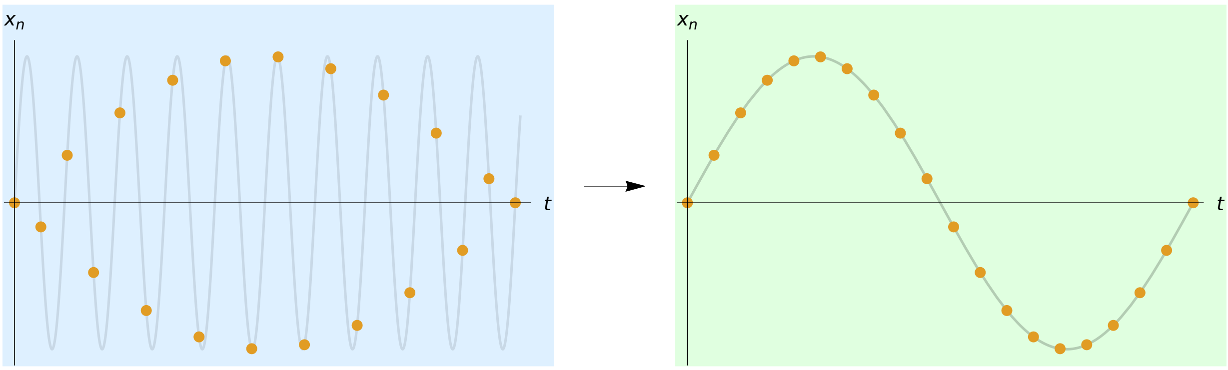

Figure 1: Schematic of the sample-rearranging algorithm. When applied to an artificially generated low-sampled data set of a sine wave (blue background), we simulate a scenario with a higher sampling rate over a single period (green background).

Moreover, it is important to note that only two requirements are necessary for a low-sampled data set to be eligible to go through the rearranging process. Its $i)$ behavior must represent a periodic regime, and $ii)$ the sampling rate must be such that we can observe a closed curve from the topology produced between the coordinates of the state vector in each 2D phase portrait.

3. RESULTS

In order to verify the performance of the sample-rearranging technique we use the Lorenz System (Sparrow, 2012) as a reference model. It takes the form

$$\dot{x}(t) = \sigma\,(y(t) - x(t)) \ , \quad \dot{y}(t) = \ x(t)\,(\rho - z(t)) - y(t) \quad \text{and} \quad \dot{z}(t) = \ x(t)y(t) - \beta z(t) \ , \tag{5}$$

where σ, ρ and β are the model parameters. As we mentioned before, our tool only works for periodic regimes. Therefore, we set $\sigma = 10,\ \beta = 8/3$ and $\rho = 400$, which asymptotically develops a periodic nonlinear regime. In this case, the initial conditions consist in a small deviation of magnitude $\delta = 10^{-5}$ from a set of the steady-states $\{x_s, y_s, z_s\}$

$$x(0) = x_s(1+\delta), \ \ y(0) = y_s(1+\delta) \ \ \text{and} \ \ z(0) = z_s(1+\delta) \ \ , \tag{6}$$

$$\text{where} \ \ x_s = \sqrt{\beta}\sqrt{\rho - 1}, \ \ y_s = \sqrt{\beta}\sqrt{\rho - 1} \ \ \text{and} \ \ z_s = \rho - 1 \ \ . \tag{7}$$

First, we produce a numerical solution of the ODE system over $t_0 = 0$ and $t_f = 200$ with 40 digits of working precision, using the function **NDSolve** from *Wolfram Mathematica* version 12.0. Figure 2 shows the 2D phase portraits after the nonlinear periodic regime development. From the numerical solution of both x and y coordinates, we directly obtain the dynamics characteristic period $p_c = 0.3630819576$. On the other hand, a cycle on $z(t)$ finishes twice as fast, then it presents a period of half magnitude instead. While this information does not affect our definition of the characteristic period, since the longer one is not only required to produce the topologies exhibited in Fig. 2 but also still fulfills the objective of the sample-rearranging technique for the $z-$coordinate, its value is highlighted by assisting us to propose an improved model when attempting to fit the time series of each coordinate.

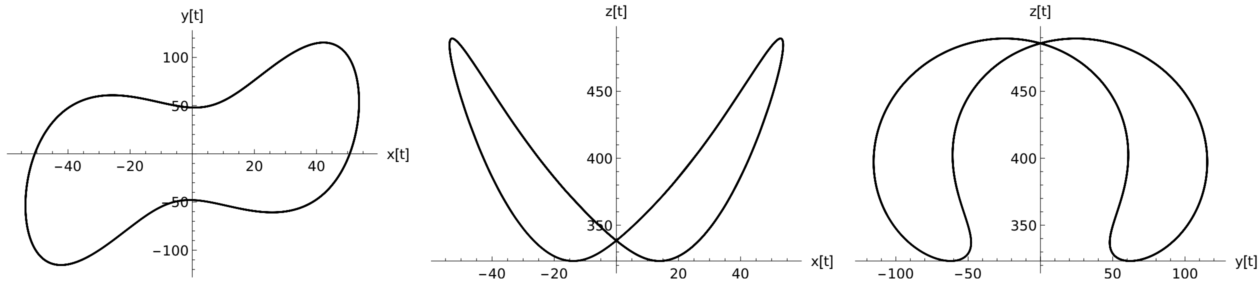

Figure 2: Periodic Lorenz system characteristic 2D phase portraits.

Our next step is artificially generating a low-sampled data set. For that, we collect 274 points employing a sampling rate of 2.735 points per period over 100 characteristic periods. Figure 3 shows the temporal behavior of the low-sampled data, in which any pattern from the original behavior is hidden in an almost chaotic distribution of points. Hence, identifying particular characteristics and proposing a model to fit the data of any of the coordinates would be a challenging process. On the other hand, when displayed in the form of 2D phase portraits, the same set outlines the topology presented by the continuous one in Fig. 2. Therefore, we can assume that most of the dynamics, even those associated with high-order nonlinearities, are present and that we dispose of the necessary information to reproduce the behavior of the original Lorenz. This contradiction exemplifies the inspiration for the development of this technique.

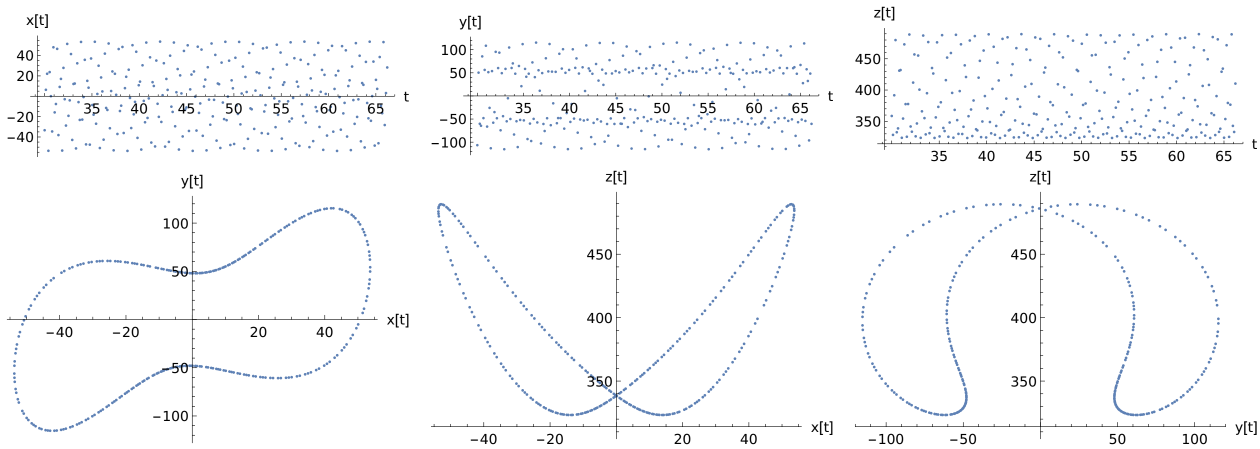

Figure 3: Temporal behavior (top) and 2D phase portraits (bottom) of the low-sampled generated data.

Finally, we rearrange the low-sampled data set based on p_c. The results can be seen in Fig. 4, in which the temporal behavior displays a completely different condition from the prior observations, simulating a scenario built by an extremely high sampling rate over a single period. Previously hidden in a cloud of points, the rearranged data set exhibits all the nonlinear dynamics revealed before only by the 2D phase portraits. Consequently, due to the periodic nature of the phenomenon, all the necessary information to properly propose a fitting model that better describes it is available.

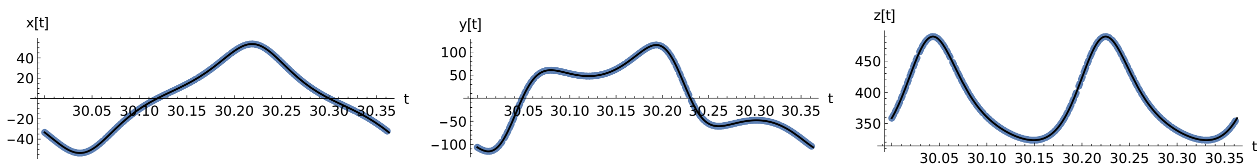

Figure 4: Temporal behavior over a single characteristic period comparison: Low-sampled data after it was rearranged (blue points) and original numerical solution (black line).

To approximate the temporal behavior of each coordinate, we perform a nonlinear regression on the rearranged data, evaluating the performance of different partial sums of the following set of infinite series as model functions

$$x(t) = \sum_{n=0}^{\infty} \alpha_n \sin(2\pi t/p_c + \phi_{x,n})^{2n+1} \quad , \quad y(t) = \sum_{n=0}^{\infty} \beta_n \sin(2\pi t/p_c + \phi_{y,n})^{2n+1} \quad \text{and}$$
$$z(t) = \gamma_0 + \sum_{n=1}^{\infty} \gamma_n \cos(4\pi t/p_c + \phi_{z,n})^{n} \quad , \tag{8}$$

where α_n, β_n, γ_0, γ_n, $\phi_{x,n}$, $\phi_{y,n}$ and $\phi_{z,n}$ are the model parameters. A relative error analysis, using the original numerical solution as a reference, shows that we achieve a more than satisfactory accuracy when the model functions are the infinite series truncated at $n = 14$.

Finally, with the approximation functions, we can proceed and construct all SINDy necessary matrices, including the challenging data matrix of derivatives in Eq. (2). By doing so, the SINDy algorithm successfully recovers the periodic Lorenz system. Figure 5 summarizes the results found, where we plot the 2D phase portraits comparing our initial low-sampled data and the solution, over the original time domain, of the ROM generated by SINDy after the sample-rearranging process based on the characteristic period.

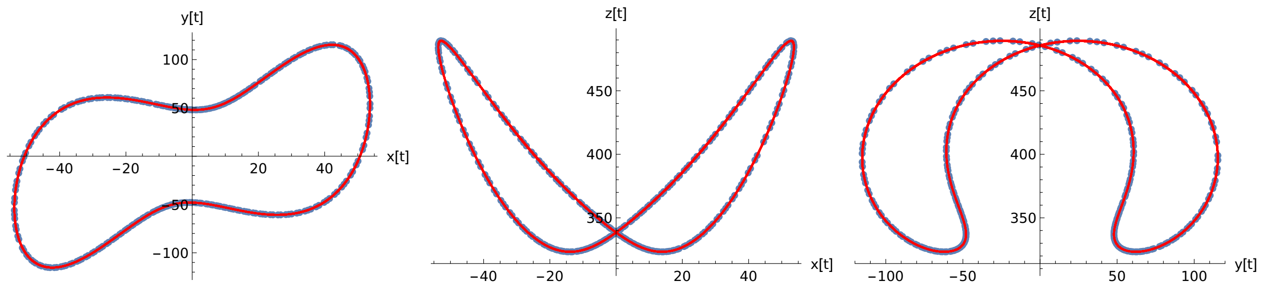

Figure 5: 2D phase portraits comparing the initial low-sampled data (blue points) and the solution of the SINDy generated ROM over 100 periods (red line).

4. CONCLUSION

In this work, we presented a new method to overcome one of the most common difficulties when developing a model for a nonlinear behavior with a data set obtained experimentally. The sample-rearranging technique takes advantage of the dynamics' periodicity to reorganize an initially low-sampled data set into one artificially taken at a high sampling rate, capturing all the information carried on 2D phase portraits closed curves topology. Here, the new technique helped SINDy to recover the Lorenz system under periodic parametric conditions.

Further evaluation of the tool is in progress, with defined future objectives. Our first intention is to test it in a more complex nonlinear scenario, such as the same Lorenz system, but under double periodic conditions instead. Finally, we also look to assess the robustness of the tool by applying it to cases of low-sampled data sets of diffusion jet flame dynamics obtained experimentally. In this case, in addition to the difficulty that naturally arises when working with noise data, precisely identifying the absent characteristic period poses another challenge.

5. ACKNOWLEDGEMENTS

The authors appreciate the financial support from CAPES, ARO, SOARD/AFOSR and ISEP/AFOSR.

6. REFERENCES

Bennewitz, J., Valentini, D., Plascencia, M., Vargas, A., Sim, H.S., Lopez, B., Smith, O. and Karagozian, A., 2018. "Periodic partial extinction in acoustically coupled fuel droplet combustion". *Combustion and Flame*, Vol. 189, pp. 46–61. doi:10.1016/j.combustflame.2017.10.019.

Bongard, J. and Lipson, H., 2007. "Automated reverse engineering of nonlinear dynamical systems". *Proceedings of the National Academy of Sciences*, Vol. 104, No. 24, pp. 9943–9948.

Boninsegna, L., Nüske, F. and Clementi, C., 2018. "Sparse learning of stochastic dynamical equations". *The Journal of chemical physics*, Vol. 148, No. 24, p. 241723.

Brunton, S.L. and Kutz, J.N., 2022. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.

Brunton, S.L., Noack, B.R. and Koumoutsakos, P., 2020. "Machine learning for fluid mechanics". *Annual review of fluid mechanics*, Vol. 52, pp. 477–508.

Brunton, S.L., Proctor, J.L. and Kutz, J.N., 2016. "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.

Champion, K., Zheng, P., Aravkin, A.Y., Brunton, S.L. and Kutz, J.N., 2020. "A unified sparse optimization framework to learn parsimonious physics-informed models from data". *IEEE Access*, Vol. 8, pp. 169259–169271. doi: 10.1109/ACCESS.2020.3023625.

Holmes, P., Lumley, J.L. and Berkooz, G., 1996. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge Monographs on Mechanics. Cambridge University Press. doi:10.1017/CBO9780511622700.

Kaheman, K., Kutz, J.N. and Brunton, S.L., 2020. "Sindy-pi: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics". *Proceedings of the Royal Society A*, Vol. 476, No. 2242, p. 20200279.

Karagozian, A., 2016. "Acoustically coupled combustion of liquid fuel droplets". *Applied Mechanics Reviews*, Vol. 68. doi:10.1115/1.4033792.

Koza, J.R., 1992. *Genetic programming: on the programming of computers by means of natural selection*, Vol. 1. MIT press.

Lettieri, D., Assis, J. and Alves, L., 2022. "Modeling single jet diffusion flame oscillations based on under-sampled data". *19th Brazilian Congress of Thermal Sciences and Engineering*.

Loiseau, J.C. and Brunton, S.L., 2018. "Constrained sparse galerkin regression". *Journal of Fluid Mechanics*, Vol. 838, pp. 42–67.

Mangan, N.M., Brunton, S.L., Proctor, J.L. and Kutz, J.N., 2016. "Inferring biological networks by sparse identification of nonlinear dynamics". *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, Vol. 2, No. 1, pp. 52–63.

Nyquist, H., 1928. "Certain topics in telegraph transmission theory". *Transactions of the American Institute of Electrical Engineers*, Vol. 47, No. 2, pp. 617–644. doi:10.1109/T-AIEE.1928.5055024.

Rudy, S.H., Brunton, S.L., Proctor, J.L. and Kutz, J.N., 2017. "Data-driven discovery of partial differential equations". *Science Advances*, Vol. 3, No. 4, p. e1602614. doi:10.1126/sciadv.1602614. URL `https://www.science.org/doi/abs/10.1126/sciadv.1602614`.

Schaeffer, H., 2017. "Learning partial differential equations via data discovery and sparse optimization". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 473, No. 2197, p. 20160446.

Schmidt, M. and Lipson, H., 2009. "Distilling free-form natural laws from experimental data". *science*, Vol. 324, No. 5923, pp. 81–85.

Sevilla-Esparza, C., Wegener, J., Teshome, S., Rodriguez, J., Smith, O. and Karagozian, A., 2014. "Droplet combustion in the presence of acoustic excitation". *Combustion and Flame*, Vol. 161. doi:10.1016/j.combustflame.2013.12.012.

Shannon, C.E., 1948. "A mathematical theory of communication". *The Bell System Technical Journal*, Vol. 27, No. 3, pp. 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Sim, H.S., Vargas, A., Ahn, D.D. and Karagozian, A.R., 2020. "Laminar microjet diffusion flame response to transverse acoustic excitation". *Combustion Science and Technology*, Vol. 192, No. 7, pp. 1292–1319. doi: 10.1080/00102202.2020.1726897.

Sirovich, L., 1987. "Turbulence and the dynamics of coherent structures part i: Coherent structures". *Quarterly of Applied Mathematics*, Vol. 45, No. 3, pp. 561–571. ISSN 0033569X, 15524485. URL `http://www.jstor.org/stable/43637457`.

Sparrow, C., 2012. *The Lorenz equations: bifurcations, chaos, and strange attractors*, Vol. 41. Springer Science & Business Media.

Tibshirani, R., 1996. "Regression shrinkage and selection via the lasso". *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 58, No. 1, pp. 267–288.

Vargas, A., 2022 - 2022. "Acoustically coupled, non-premixed combustion processes / by andres vargas."

Vargas, A., Guerrero, J., Kiani, S. and Karagozian, A., 2020. "Laminar Flame Dynamics of Multi-Port Fuel Jets Under Acoustic Forcing". In *APS Division of Fluid Dynamics Meeting Abstracts*. APS Meeting Abstracts, p. Z12.002.

Zhang, S. and Lin, G., 2018. "Robust data-driven discovery of governing physical laws with error bars". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, Vol. 474, p. 20180305. doi: 10.1098/rspa.2018.0305.

Zheng, P., Askham, T., Brunton, S.L., Kutz, J.N. and Aravkin, A.Y., 2018. "A unified framework for sparse relaxed regularized regression: Sr3". *IEEE Access*, Vol. 7, pp. 1404–1423.

7. RESPONSIBILITY NOTICE

The authors are solely responsible for the printed material included in this paper.