

MPhil DIS Project 24

Executive Summary



CRSiD: tmb76

Department of Physics

University of Cambridge

*Submitted in partial fulfilment of the requirements of the MPhil degree in Data
Intensive Science*

Hughes Hall

June 30, 2024

Contents

0.1	Introduction (100 words)	1
0.2	Methodology (250 words)	2
0.3	Conducted Research and Results (350 words)	3
0.4	Application to Elasto-Inertial Turbulence (100 Words)	3
0.5	Conclusion (100 Words)	4

0.1 Introduction (100 words)

In a majority of simulation problems in physics and engineering, it is usually unnecessary to consider the full system of equations that govern it. Rather, it is often possible to simplify the equations by considering a subset of the terms to be negligible compared to the others. And because the other terms dominate the dynamics of the system, simulating it with only these dominant terms can provide a good approximation, whilst being computationally cheaper. This concept is the fundamental idea behind the method of dominant balance, where the subset of dominant terms is assumed to be in balance. And this method has proven extremely useful in multiple fields, including meteorology [2, 4, 7, 10].

Historically, this method has been applied manually by field experts, often over a long period of researching and the manipulation of complex mathematical equations. Over the recent years, there has been a growing interest in automating this process using machine learning techniques. Most of these however, were limited to specific systems and sometimes relied on expert interpretation rather than automatic identification of the dominant balances [?, 6, 9]. A similar endeavour has also been pursued, with the application of machine learning techniques in directly identifying the governing equation of a physical system from simulation data [1, 5].

In this project, a novel method proposed by Callaham et al. (2021) [3] is explored, verified, and used. This method proposes a highly generalisable method to identify dominant balance models in a variety of physical systems, with minimal input from

the user. However, before using it for new systems, it is important to verify its robustness and reproducibility. One of this project's core aims is to try to reproduce the results of the original paper, by using alternative code. This is to ensure that the paper's results are not dependent on the specific implementation of the method, but rather the method itself. In terms of coding, this means testing whether there has been no lucky seeding or bug in the original code that has led to the results. Additionally, the choice of clustering algorithm is discussed, and the stability of the method under change of hyperparameters is explored. The method is then applied to simulation data of a fairly recent flow called elasto-inertial turbulence, to demonstrate its potential in uncovering new dominant balance regimes in complex flows [8]

0.2 Methodology (250 words)

The proposed method can be summarized into 3 steps. The first step involves transforming the data from the physical system into an equation-space format. This is done by computing the terms of the governing equations of the system, at each point in the physical space. Then, by considering each point in physical space as a sample with features the values of the equation's terms. The second step then makes full use of this new format by clustering the samples in equation space, therefore grouping points with terms of similar magnitude. This is done using a Gaussian Mixture Model clustering algorithm, which has the advantage of dealing very well with clusters of different shapes and sizes, and only requires for the number of clusters to be set. The final step involves applying Sparse Principal Component Analysis (SPCA) to the clustered data. SPCA is a variant of PCA where a regularization constraint is applied to the number of non-zero coefficients in the principal components. Essentially, instead of returning a full principal component that is always a linear combination of all the features, SPCA returns a sparse principal component that is a linear combination of only a few features, with some of the coefficients being zero. By applying this to the points in each cluster, and only taking the leading principal component, this step returns a vector which determines what subset of the terms of the governing equation are in balance in that cluster (0 if the term is negligible, non-zero if it is dominant). The severity with which SPCA penalizes the number of non-zero terms is determined by the second hyperparameter of this method; α . The

larger the value, the higher the chance of a term being considered inactive. In the case where step 2 identified a too large number of cluster, step 3 has the particularity of identifying clusters of points that have the same balance model.

0.3 Conducted Research and Results (350 words)

Discuss the portability and reproducibility of the code, including any issues encountered and resolved. Detail the turbulent boundary layer case: How the original and alternative codes were used to reproduce the results. Key findings from the reproduction of the RANS equation's terms and GMM clustering. Results from applying SPCA to identify active terms. Summarize the exploration of other algorithms: Results from Spectral Clustering, K-Means, and Weighted K-Means. Highlight the stability assessment: Impact of different numbers of clusters and alpha values on the results. Effect of training set size on the identified balance models.

Figures to include:

Figure 5.1: Plot of the 6 terms in the RANS equation using the original Callaham code. Figure 5.2: Plot of the 6 terms in the RANS equation using alternative code. Figure 5.3: Covariance matrices for each cluster found by the sklearn and custom GMM algorithms. Figure 5.4: Plot of the clusters found by the sklearn and custom GMM algorithms. Figure 5.5: Plot of unique balance models found after applying SPCA with the original and alternative code. Figure 5.6: Plot of unique balance model clusters in physical space with the original and alternative code.

0.4 Application to Elasto-Inertial Turbulence (100 Words)

Briefly describe the application of the method to a new dataset involving elasto-inertial turbulence. Highlight the potential of the method to uncover new dominant balance regimes in complex flows.

Figures to include:

Figure 6.1: Plot of the attractors found in the DNS of the FENE-P model.

0.5 Conclusion (100 Words)

Summarize the key findings and their implications. Reflect on the strengths and limitations of the method. Suggest potential future work or improvements.

Bibliography

- [1] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data: Sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.
- [2] A. P. Burger. Scale consideration of planetary motions of the atmosphere. *Tellus*, 10(2):195–205, 1958.
- [3] J.L. Callahan, J.V. Koch, and B.W. et al. Brunton. Learning dominant physical processes with data-driven balance models. *Nature Communications*, 12:1016, 2021.
- [4] J. G. Charney. The dynamics of long waves in a baroclinic westerly current. *Journal of Atmospheric Sciences*, 4:136–162, 1947.
- [5] Miles Cranmer, Alvaro Sanchez-Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases. *arXiv preprint arXiv:2006.11287*, 2020.
- [6] Jin Lee and Tamer A. Zaki. Detection algorithm for turbulent interfaces and large-scale structures in intermittent flows. *Computers & Fluids*, 175:142–158, 2018.
- [7] Norman A. Phillips. Geostrophic motion. *Reviews of Geophysics*, 1:123–176, 1963.
- [8] Devranjan Samanta, Yves Dubief, Markus Holzner, Christof Schäfer, Alexander Morozov, Christian Wagner, and Björn Hof. Elasto-inertial turbulence. *Proceedings of the National Academy of Sciences*, 110:10557 – 10562, 2012.
- [9] Matthias Sonnewald, Carl Wunsch, and Patrick Heimbach. Unsupervised learning reveals geography of global ocean dynamical regions. *Earth and Space Science*, 6:784–794, 2019.

- [10] Jun-Ichi Yano and M. Bonazzola. Scale analysis for large-scale tropical atmospheric dynamics. *Journal of Atmospheric Sciences*, 66:159–172, 2009.