49th Leeds-Lyon Symposium on Tribology – September 2-4, 2024, Lyon, France
50 years of Leeds-Lyon

DATA REDUCTION TECHNIQUES FOR OPTIMISED MACHINE LEARNING MODELLING OF TEHL CONTACTS IN JOURNAL BEARINGS

S Cartwright [a], K Elkobrossy [b], B C Rothwell [a*], G Figueredo [b], H Medina [a], C Eastwick [a], S Ambrose [a], J Layton [a]
*Benjamin.rothwell@nottingham.ac.uk
[a] Mechanical and Aerospace Systems,
Faculty of Engineering, University of Nottingham, Nottingham, NG7 2RD, United Kingdom

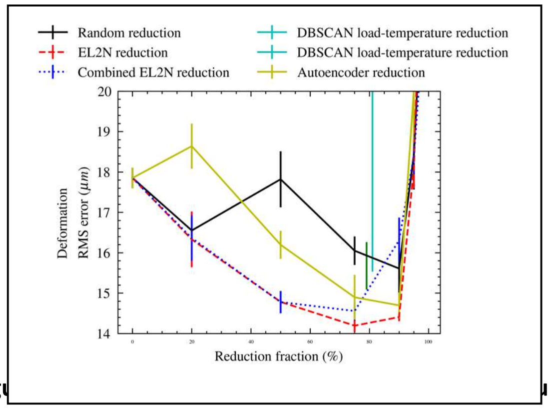
[b] School of Computer Science,
University of Nottingham, Nottingham, NG7 2RD, United Kingdom

KEYWORDS
Machine learning / AI; Hydrodynamic Lubrication; Modelling in tribology, Data reduction

ABSTRACT
A common problem faced in the development of Machine Learning (ML) systems is the computational complexity of training models such as neural networks. This effect is compounded as the size of both the model and the training dataset increase. Previous work has resulted in the development of several algorithms to reduce the size of training datasets by selecting for the most important data points and filtering out the least.

While numerous applications for ML models have been found for the simulation of tribological contacts, the cost of training models remains high especially when maximising dataset size for optimum performance. By applying data reduction algorithms to a neural network-based Thermoelastic Hydrodynamic Lubrication dataset of $3.9 * 10^7$ points [1], superior performance can be obtained using the same amount of computation applied to an optimised training dataset where data points deemed less useful for training the model are removed. The result of this is that the break-even point at which the computational cost of developing an ML solution is less than the use of a conventional numerical model is reached earlier.

In this study, four data reduction algorithms are implemented (random reduction, error L2 norm, density-based spatial clustering and a regression autoencoder). Datasets with reduction rates of 20, 50, 75, 90, 95 and 99% are generated. The distribution of the optimized datasets is analysed to provide insight into how data should be selected to improve model performance. The reduced datasets are used to train neural network models and their performance is evaluated against a full-sized and randomly reduced dataset for predicting several journal bearing performance characteristics.



Figu... urface deformation for different size datasets produced using various algorithms.

Evaluation is carried out after a fixed training period of $5*10^6$ iterations to study the relationship between performance and dataset optimization with a constant computational cost.

At extreme reduction rates of >95%, the tested algorithms tend to perform worse than randomly sampling data due to low priority regions of the dataset being underrepresented. For more moderate reduction rates, between 20 and 75%, the algorithmically reduced datasets frequently produce models with mean squared errors over 10% less than with randomly sampled datasets. This work demonstrates the potential for optimised datasets to improve the performance of ML-based models for tribological processes without increasing dataset size or model complexity.

REFERENCES
1. Cartwright, S., et al., *A Machine Learning-Driven Approach to predicting Thermo-Elasto-Hydrodynamic Lubrication in Journal Bearings.* Tribology International, 2023.