Using Artificial Neural Networks to Accelerate Thermo-elastohydrodynamic Lubrication Simulations

Filimonas Kaliafetis[a,*], Suhaib Ardah[a], James P. Ewen[a], Daniele Dini[a]

[a]*Department of Mechanical Engineering, Imperial College London, Exhibition Road, South Kensington, London, SW7 2AZ, United Kingdom*

Abstract

Thermo-elastohydrodynamic lubrication (TEHL) simulations have long been used to predict the lubrication performance of various engineering systems. The aim of this study is to test the performance of artificial neural networks (ANNs) in predicting important parameters in tribological systems and to test the speed of hybrid simulation frameworks that combine ANNs with classical finite volume method (FVM) approaches to solve TEHL problems for point contacts. Six different ANNs were trained to predict values of rigid separation between the contacting bodies, maximum fluid temperature, mid-film friction coefficient, as well as profiles of hydrodynamic pressure, liquid film friction and fluid temperature. In total, 600 classical TEHL simulations were used as training and testing data for the ANNs. The ANN predictions were very accurate. The rescaled R^2 values (in the range of 0 to 1), indicating the correlation between the ANN predictions and classical TEHL results, were 0.997 for the rigid separation, 0.985 for the maximum fluid temperature and 0.939 for the friction coefficient. The mean rescaled R^2 values across all testing samples were 0.997 for both pressure and liquid film fraction profiles and 0.873 for the fluid temperature profiles. Using ANNs to initialise the rigid separation, pressure and liquid film fraction profiles and the temperature profile in each simulation results to significant reduction in simulation times with the mean simulation time across all testing samples being about 40% less than the mean time of the classical TEHL simulations, and for individual simulations, the hybrid frameworks being up to three times faster.

Keywords: Thermo-elastohydrodynamic lubrication, machine learning, artificial neural networks, predictive lubrication modelling

1. Introduction

Numerical simulations have long been used as a tool for predicting the performance of various engineering systems. In the field of tribology, many techniques have been used to predict lubrication performance, surface evolution and component reliability [1, 2].

One specific approach involves the use of elastohydrodynamic lubrication (EHL) simulations, which solve the Reynolds equation in an iterative manner to calculate pressure and film thickness profiles for a viscous lubricant film separating two solid bodies. Elastic deformation of the solid bodies and viscosity increase of the lubricant due to the applied pressure are taken into account in addition to lubricant starvation, non-Newtonian behaviours, transient effects and the presence of surface coatings [3].

Thermo-elastohydrodynamic lubrication (TEHL) simulations are a special case of EHL simulations in which the energy equations for the lubricant film and solid bodies are solved alongside the Reynolds equation in a coupled manner to calculate the temperature distributions in the lubricant film and solids in addition to the pressure and film thickness profiles. Some important TEHL studies include those of Cheng [4], who compared the simulation results for a line contact with the experimental data of Crook [5]. The film thickness results of the TEHL simulations were in good agreement with the experimental data. Zhu and Wen [6] presented a full, steady-state numerical solution for elliptical TEHL contacts, calculating pressure, film thickness as well as the three-dimensional temperature distribution in the lubricant film and solids. Liu et al. [7] developed a non-Newtonian numerical solution system for TEHL problems and proposed a novel technique to determine the effective lubricant viscosity. Habchi et al. [8] obtained a complete solution of the energy equation applied to both contacting solids and the lubricant film with the use of complex rheological equations for shear thinning and temperature effects. They compared calculations for film thickness and friction with experimental measurements, showing good agreement between the two. The study also showed that the TEHL simulation results were closer to reality than the results from isothermal simulations. More recently, Wang et al. [9] developed a unified TEHL model capable of simulating the entire lubrication regime from full-film lubrication down to boundary lubrication, while also taking into account the effects of surface roughness.

As the complexity of TEHL simulations has been increasing over the years, considerable efforts have been made to reduce computational times through the use of efficient algorithms like Fast Fourier Transforms (FFT) for calculation of surface deformations [10], techniques like progressive mesh densification [11] as well as relaxation schemes like the Point Gauss-Seidel

*Corresponding author

Email address: f.kaliafetis23@imperial.ac.uk (Filimonas Kaliafetis)

with Aitken Acceleration (PGMA) scheme all of which speed up numerical convergence of TEHL solvers [12]. Despite the use of the aforementioned methods, TEHL simulations still require significant computational resources and take a considerable amount of time to reach a converged solution. For example, Lohner et al. [13] mentioned that one particular set of steady state TEHL simulations for counterformal line contacts converged in approximately 6 to 8 minutes on a computer with a 3.5 GHz processor, while Ardah et al. [12] reported simulation times for line contacts ranging from 5 minutes to approximately 40 hours depending on the mesh resolution and parameters used. For three-dimensional problems simulation times are even longer.

More recently, machine learning (ML) and artificial intelligence (AI) have emerged as powerful engineering tools because they allow prediction of desired parameters faster than classical simulations even in complex systems [14]. In tribology, various ML and AI approaches have been used to predict friction and wear as well as classify equipment life and forecast component failure [15, 16, 17, 18, 19]. In the context of EHL simulations, Marian et al. [20] used different ML algorithms to predict the central and minimum film thickness of lubricant films for isothermal line and point contacts using training data from EHL simulations, showing that ML approaches can predict film thickness parameters more accurately than well-established empirical equations and faster than classical EHL simulations. Habchi and Bair [21] also used ML to predict the minimum film thickness of lubricant films in isothermal circular contacts, again using data from EHL simulations for training. Tošić et al. [22] conducted a similar study in which different Artificial Neural Networks (ANNs) were used to predict EHL film thickness in elliptical contacts with varying direction of lubricant entrainment. Their results showed that ANNs can successfully be trained to predict film thickness parameters over a wide range of operating conditions. Their approach also highlighted some good practices for constructing ML models, examining the selection of ML algorithms, selection of ANN hyperparameters, as well as the choice and size of training data. Walker et al. [23] also used ANNs to predict EHL central film thickness, viscous and boundary friction, in multiple commonly used machine elements, such as a rolling element bearing and a spur gear. The viscous and boundary friction ANN predictions were validated against ball-on-disc experimental measurements [23]. Issa et al. [24] used Gaussian Process Regression (GPR), another ML tool, to predict the central and minimum film thickness for isothermal elliptical contacts with a wide range of ellipticity and operating conditions. The GPR predictions were found to be more accurate than conventional film thickness analytical formulae. Kelley et al. [25] trained an ANN to evaluate the full solution to the Reynolds equation in the form of predicted pressure and film thickness profiles for isothermal line contacts. Their results showed that ANN predictions were more accurate than classical curve-fitting models. Finally, Singh et al. [26] used neural networks (NNs) to predict the maximum temperature in line contacts. Results from TEHL simulations were used for the training of the NNs.

In the above studies, the ML models are used as surrogate models to replace the classical simulations for predicting film thickness or temperature parameters. Despite their very good accuracy, the predicted ML results can only serve as an approximation to the results obtained from the classical EHL simulations. Moreover, in the study of Habchi and Bair [21], the study of Tošić et al. [22] and the study of Issa et al. [24] the predicted parameters are scalar values of minimum and/or central film thickness and not detailed film thickness profiles. This is also the case in the study of Singh et al. [26], in which the predicted outputs are scalar values of maximum temperature and not detailed temperature profiles. While Kelley et al. [25] predicted detailed pressure and film thickness profiles, there is a need to expand their work to non-isothermal point contacts.

In summary, classical TEHL simulations have been used to accurately model the behaviour of complex tribological systems. However, because of the complexity of the simulations and high computational requirements, they often take a considerable amount of time to converge. On the other hand, ML models can predict the behaviour of tribological systems in fractions of a second once trained, but do not offer the same level of accuracy as TEHL simulations. Hence, there is a need to combine the accuracy of TEHL simulations with the speed of ML models.

The aim of this study is to test the performance of ML models in predicting important parameters in tribological systems as well as to test the speed of hybrid frameworks that combine ML models with classical finite volume approaches to solve TEHL problems for point contacts.

2. Methodology

This section outlines the governing equations and numerical implementation underlying the TEHL framework. This is followed by a description of the hybrid modelling strategy designed to accelerate the TEHL simulations by leveraging neural network predictions as initial approximations. The methodology for generating the training and testing datasets is described in detail, along with a summary of the simulation parameters utilised. Subsequently, the architecture and training of the ML models employed in this study are presented. The three-dimensional point contact geometry considered in the current contribution is depicted in Figure 1.

2.1. Thermo-elastohydrodynamic Lubrication Framework

The thermo-elasohydrodynamic lubrication (TEHL) model employed in this study builds upon the computational framework introduced by Ardah et al. [27]. Within this framework, the governing equations are discretised using the Finite Volume Method (FVM) and solved over three-dimensional curvilinear grids defined in the computational space. An overview of the simulation framework is portrayed in Figure 2, while the formulation of the governing equations is detailed in the subsequent sections.

As illustrated in Figure 2, the TEHL simulation follows an iterative procedure that requires initial estimates for the rigid separation (h_0) between the two solid bodies as well as initial

approximations for the pressure (p) and temperature (T) distributions. The initial pressure profile is based on the classical Hertzian contact solution, while the temperature is assumed to be uniform and equal to the ambient value throughout the computational domain [12]. The initial estimate for the rigid body separation is computed using the Hamrock and Dowson central film thickness equation for point contacts, as provided below [28]:

$$h_{0,HD} = 2.69RU^{*0.67}G^{*0.53}W^{*-0.067}(1-0.61e^{-0.73}), \quad (1)$$

where

$$U^* = \frac{\eta_0 u_{ent}}{E'R}, \quad G^* = \alpha_p E', \quad W^* = \frac{W}{E'R^2}.$$

The isothermal EHL problem is solved first to find the converged pressure and film thickness distributions that balance the applied load. In each EHL iteration, the generalised Reynolds equation (Eq. (3)) and the elastic deformation equation (Eq. (6)) are solved, while also updating the film thickness geometry (Eq. (5)) and lubricant properties (Eqs. (12)-(14)). This is repeated until the combined pressure and liquid film fraction residue ($r^{p,\theta}$) is smaller than the combined pressure-liquid film fraction convergence criterion ($e^{p,\theta}$).

Once load convergence has been reached, indicated by the load residue (r^W) being smaller than the load convergence criterion (e^W), the thermo-EHL problem can be solved. This is done iteratively by calculating the lubricant and solid surface temperature distributions for the converged pressure and film thickness solutions until the temperature residue (r^T) is smaller than the temperature convergence criterion (e^T). Once temperature convergence has been reached, the global pressure and temperature convergence is checked. If both the global pressure (r^P_{global}) and temperature (r^T_{global}) terms are smaller than the global pressure and temperature convergence criteria (e^p_{global} and e^T_{global}) then a global converged solution has been found and the simulation can end.

The residue terms mentioned above are defined as follows:

$$r^{p,\theta} = \frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}|p^n_{i,j}-p^{n-1}_{i,j}|}{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}|p^n_{i,j}|} + \frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}|\theta^n_{i,j}-\theta^{n-1}_{i,j}|}{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}|\theta^n_{i,j}|}, \quad (2a)$$

$$r^W = \frac{\left|W - \iint_\Omega p(x,y)\,dx\,dy\right|}{W}, \quad (2b)$$

$$r^T = \frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_z}|T^n_{i,j,k}-T^{n-1}_{i,j,k}|}{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_z}|T^n_{i,j,k}|}, \quad (2c)$$

$$r^p_{global} = \frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}|p^g_{i,j}-p^{g-1}_{i,j}|}{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}|p^g_{i,j}|}, \quad (2d)$$

$$r^T_{global} = \frac{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_z}|T^g_{i,j,k}-T^{g-1}_{i,j,k}|}{\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_z}|T^g_{i,j,k}|}. \quad (2e)$$

where i, j and k refer to the x, y and z nodes in the domain respectively, N_x, N_y and N_z refer to the mesh size in the x-, y- and z-directions respectively, n refers to the current EHL or TEHL iteration and g refers to the current global iteration.

2.1.1. Hydrodynamic Flow

Considering the surface contact kinematics defined by the velocity fields $\mathbf{v_1}(x,y,t) = \begin{bmatrix} u_1 & v_1 \end{bmatrix}^T$ and $\mathbf{v_2}(x,y,t) = \begin{bmatrix} u_2 & v_2 \end{bmatrix}^T$ in the x- and y-directions, along with the spatiotemporal variations in lubricant film thickness $h(x,y,t)$, dynamic viscosity $\eta(x,y,t)$ and density $\rho(x,y,t)$, the hydrodynamic pressure distribution within the lubricant film is governed by the *generalised* Reynolds equation. Furthermore, the mass-conserving $p-\theta$ Elrod-Adams cavitation model is employed to ensure the so-called JFO complementary boundary conditions in order to account for the fluid film rupture, *i.e.* cavitation. Therefore, the generalised Reynolds equations is formulated in terms of the coupled pressure-fraction film content ($p-\theta$) framework, which facilitates a numerically conservative treatment of lubricant rupture and reformation. Accordingly, the resulting $p-\theta$ generalised Reynolds equation, expressed in expanded and strongly conservative form, is given as follows:

$$\underbrace{\frac{\partial}{\partial x}\left(\epsilon\frac{\partial p}{\partial x}\right)+\frac{\partial}{\partial y}\left(\epsilon\frac{\partial p}{\partial y}\right)}_{\substack{\text{Pressure flow}\\ \text{terms}}} = \underbrace{\frac{\partial(\theta\rho^*_x)}{\partial x}+\frac{\partial(\theta\rho^*_y)}{\partial y}}_{\substack{\text{Entraining flow}\\ \text{terms}}} + \underbrace{\frac{\partial(\theta\rho_e)}{\partial t}}_{\substack{\text{Transient}\\ \text{term}}}, \quad (3)$$

where the associated coefficients and integrals which account for the variations in the lubricant viscosity and density across the film thickness are defined as follows:

$$\epsilon = \frac{\eta_e}{\eta'_e}\rho' - \rho'', \qquad \rho^*_x = \rho_e u_1 + \eta_e u_s \rho',$$

$$\rho^*_y = \rho_e v_1 + \eta_e v_s \rho', \qquad \rho_e = \int_{z_1}^{z_2}\rho\,dz,$$

$$\frac{1}{\eta_e} = \int_{z_1}^{z_2}\frac{dz}{\eta}, \qquad \rho' = \int_{z_1}^{z_2}\rho\int_{z_1}^{z}\frac{dz'}{\eta}\,dz,$$

$$\frac{1}{\eta'_e} = \int_{z_1}^{z_2}\frac{z\,dz}{\eta}, \qquad \rho'' = \int_{z_1}^{z_2}\rho\int_{z_1}^{z}\frac{z'dz'}{\eta}\,dz.$$

The complementarity boundary conditions that govern the coexistence of the pressurised and cavitated regions within the lubricant film are expressed as:

$$(p-p_{cav})(1-\theta) = 0 \rightarrow \begin{cases} p > p_{cav} \rightarrow \theta = 1 & \text{pressured zones,} \\ p = p_{cav} \rightarrow 0 \le \theta < 1 & \text{cavitated zones,} \end{cases} \quad (4)$$

where p_{cav} is the cavitation pressure , while θ is the film fraction representing the proportion of liquid lubricant within the biphasic mixture that arises due to cavitation, defined mathematically as $\theta = \rho/\rho_c$, where ρ_c denotes the density of the fully saturated liquid lubricant at the cavitation pressure.

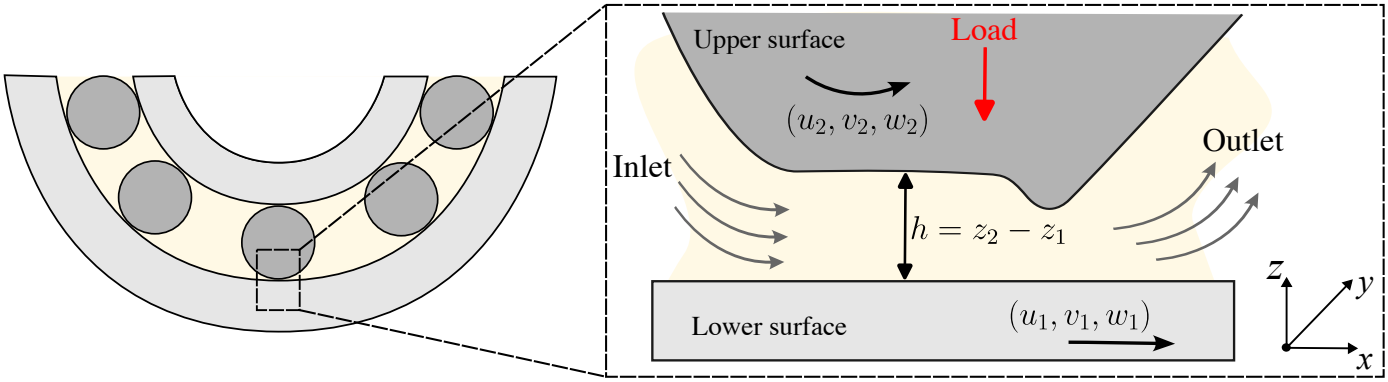


Figure 1: A schematic illustrating the three-dimensional point contact configuration.

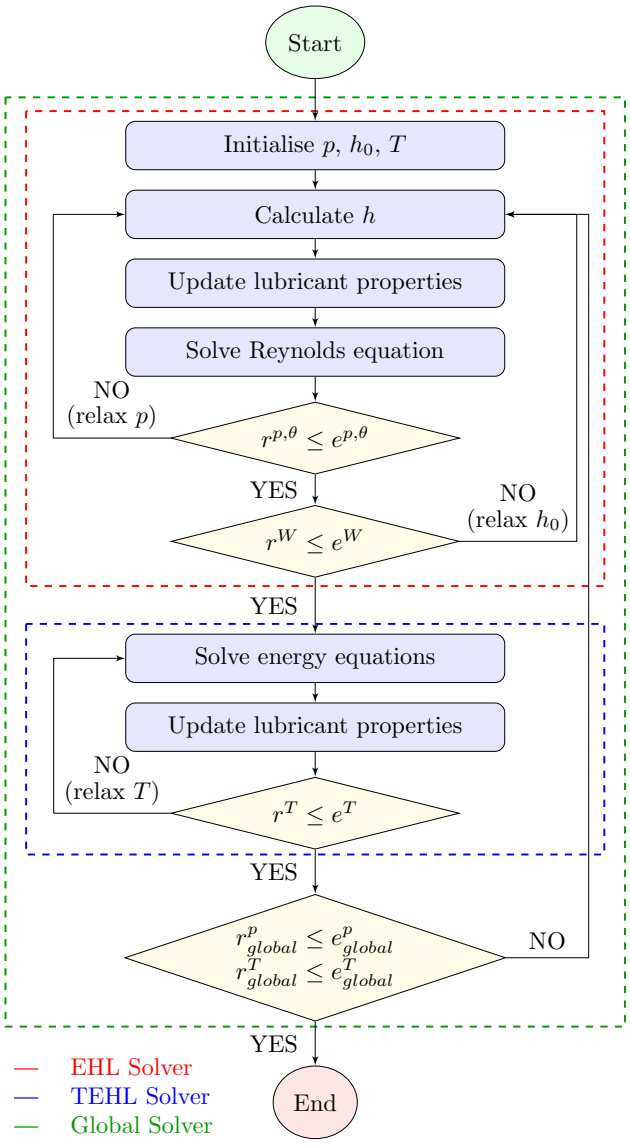


Figure 2: TEHL simulation procedure

2.1.2. Lubricant Film Geometry

The geometry of the lubricant gap in a point contact EHL can be described in Cartesian coordinates using the following expression:

$$h(x, y) = h_0 + \frac{x^2}{2R_x} + \frac{y^2}{2R_y} + \delta(x, y), \tag{5}$$

where h_0 denotes the rigid separation between the contacting bodies, R_x and R_y are the equivalent radii of curvature at the centre of contact along the x- and y-directions, respectively ($R_x = R_y = R$ for circular point contacts).

The normal elastic displacement $\delta(x, y)$ induced by the hydrodynamic pressure distribution is evaluated using the Boussinesq solution for a three-dimensional elastic half-space and is given by the following convolution integral:

$$\delta(x, y) = \frac{2}{\pi E'} \iint_{\Omega} \frac{p(\xi, \varsigma)}{\sqrt{(x - \xi)^2 + (y - \varsigma)^2}} \, d\xi \, d\varsigma, \tag{6}$$

where $p(\xi, \varsigma)$ is the pressure distribution over the contact domain Ω, and E' is the reduced (or effective) elastic modulus of the contacting materials, defined as:

$$\frac{2}{E'} = \left[\frac{(1 - \nu_1{}^2)}{E_1} + \frac{(1 - \nu_2{}^2)}{E_2} \right]^{-1},$$

where $E_{1,2}$ and $\nu_{1,2}$ are the Young's moduli and Poisson's ratios of the two contacting solids, respectively. Readers interested in the numerical evaluation of the Boussinesq solution are referred to [2] for implementation details and efficient computation strategies.

2.1.3. Lubricant Film Temperature

The temperature distribution within the lubricant film is described by the three-dimensional energy equation for compressible, viscous fluids. In its strong conservation form, and expressed in compact index notation over Cartesian coordinates

x^j with $j = \{1, 2, 3\}$, the equation reads:

$$\underbrace{\frac{\partial\left(\rho c_p T\right)}{\partial t}}_{\substack{\text{Transient}\\\text{Term}}} + \underbrace{\frac{\partial}{\partial x^j}\left(\rho c_p u_j T\right)}_{\substack{\text{Convection}\\\text{Term}}} = \underbrace{\frac{\partial}{\partial x^j}\left(k\frac{\partial T}{\partial x^j}\right)}_{\substack{\text{Conduction}\\\text{Term}}} + \underbrace{Q_T}_{\substack{\text{Source}\\\text{Term}}}, \tag{7}$$

$$Q_T = Q_p + Q_{cp} + Q_\Phi + \dot{q}_V,$$

where T is the film temperature, ρ, c_p and k are the lubricant density, specific heat capacity and thermal conductivity. The velocity components in the three spatial directions are denoted by u_j, and summation over repeated indices is implied (Einstein notation). The source term Q_T in Eq. (7) encompasses all internal heat generation mechanisms and is expressed as follows:

$$Q_p = \underbrace{\beta T\left(\frac{\partial p}{\partial t} + u\frac{\partial p}{\partial x} + v\frac{\partial p}{\partial y}\right)}_{\substack{\text{Compressive}\\\text{Heating/Cooling}}}, \tag{8a}$$

$$Q_{cp} = \underbrace{\rho T\left(\frac{\partial c_p}{\partial t} + u\frac{\partial c_p}{\partial x} + v\frac{\partial c_p}{\partial y} + w\frac{\partial c_p}{\partial z}\right)}_{\substack{\text{Enthalpic}\\\text{Heating/Cooling}}}, \tag{8b}$$

$$Q_\Phi = \underbrace{\eta\left[\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2\right]}_{\substack{\text{Shear}\\\text{Heating}}}. \tag{8c}$$

Additionally, the velocity components of the lubricant flow at the contact interface contributing to the convective heat flux are expressed as:

$$u(x,y,z) = u_1 + \frac{\partial p}{\partial x}\left(\int_{z_1}^{z}\frac{z'dz'}{\eta} - \frac{\eta_e}{\eta'_e}\int_{z_1}^{z}\frac{dz'}{\eta} + \eta_e u_s\int_{z_1}^{z}\frac{dz'}{\eta}\right), \tag{9a}$$

$$v(x,y,z) = v_1 + \frac{\partial p}{\partial y}\left(\int_{z_1}^{z}\frac{z'dz'}{\eta} - \frac{\eta_e}{\eta'_e}\int_{z_1}^{z}\frac{dz'}{\eta} + \eta_e v_s\int_{z_1}^{z}\frac{dz'}{\eta}\right), \tag{9b}$$

$$w(x,y,z) = -\frac{1}{\rho}\frac{\partial}{\partial x}\int_{z_1}^{z}\rho u\,dz' - \frac{1}{\rho}\frac{\partial}{\partial y}\int_{z_1}^{z}\rho v\,dz'. \tag{9c}$$

2.1.4. Surface Contact Temperature

Similar to the thermal evaluation of the entrained lubricant film presented in Section 2.1.3, the thermal behaviour of the bounding solids is evaluated under the assumption that internal heat generation within the solid bodies is negligible. This simplification emphasises the importance of heat conduction and convection as the dominant modes of thermal transport in the solids. Consequently, and using subscripts $s = 1$ and $s = 2$ to denote the lower and upper contacting solids, respectively, the transient temperature distribution within the solid domains is governed by the following energy conservation equation:

$$\underbrace{\frac{\partial\left(\rho_s c_{p_s} T_s\right)}{\partial t}}_{\substack{\text{Transient}\\\text{Term}}} + \underbrace{\frac{\partial}{\partial x^j}\left(\rho c_{p_s}\boldsymbol{v}_{j_s} T_s\right)}_{\substack{\text{Convection}\\\text{Term}}} = \underbrace{\frac{\partial}{\partial x^j}\left(k_s\frac{\partial T_s}{\partial x^j}\right)}_{\substack{\text{Conduction}\\\text{Term}}}, \tag{10}$$

where ρ_s, c_{p_s} and k_s represent the density, specific thermal capacity and thermal conductivity of the contacting body (s), respectively. The term $\boldsymbol{v}_{j_s}$ denotes the velocity vector of the solid's surface, defined as $\boldsymbol{v}_{j_s} = \begin{bmatrix} u_s & v_s & w_s \end{bmatrix}^T$. This velocity matrix captures the spatial motion of the solid surface in the coordinate directions.

2.1.5. Friction Coefficient

The friction coefficient (μ) is determined as the ratio of the total tangential force generated within the lubricant film to the externally applied normal load (W). This can be expressed as:

$$\mu = \frac{\int \tau\,dx\,dy}{W} = \frac{\int \eta(\partial u/\partial z)\,dx\,dy}{W}, \tag{11}$$

where τ is the shear stress within the fluid film, η is the dynamic viscosity and $\partial u/\partial z$ is the shear rate in the direction normal to the sliding surfaces. This integration is performed over the contact area in the xy plane.

2.1.6. Lubricant Viscosity

The Roelands equation is employed to evaluate the combined effects of pressure and temperature on the lubricant's dynamic viscosity under elastohydrodynamic conditions, given as [29]:

$$\eta = \eta_0 \exp\Big\{[\ln\eta_0 + 9.67]\times \Big[(1 + 5.1\times10^{-9}p)^z \times \left(\frac{T-138}{T_0-138}\right)^{-s} - 1\Big]\Big\}, \tag{12}$$

where

$$z = \frac{\alpha_p}{5.1\times10^{-9}(\ln\eta_0 + 9.67)},$$
$$s = \frac{0.0476(T_0 - 138)}{\ln\eta_0 + 9.67},$$

using the reference viscosity of the lubricant η_0, the pressure-viscosity coefficient α_p and the reference (or ambient) temperature T_0.

Furthermore, the Eyring shear thinning model is utilised to account for non-Newtonian behaviour within the lubricant as follows [30]:

$$\dot{\gamma} = \frac{\tau_E}{\eta}\sinh\left(\frac{\tau}{\tau_e}\right), \tag{13}$$

where $\dot{\gamma}$ denotes the shear rate, τ is the shear stress and τ_E is the Eyring shear stress, taken in this study to be 1×10^7 Pa. This model effectively captures the shear-thinning response observed in many lubricants under high-pressure, high-shear conditions typical of elastohydrodynamic contacts.

2.1.7. Lubricant Density

The pressure and temperature dependence of the lubricant density ρ is modelled using the Dowson-Higginson empirical relationship expressed as follows[31]:

$$\rho = \rho_0\left(1 + \frac{0.6 \times 10^{-9} p}{1 + 1.7 \times 10^{-9} p}\right)[1 - 6.5 \times 10^{-4}(T - T_0)], \quad (14)$$

where ρ_0 is the reference density corresponding to ambient pressure and temperature T_0, taken to be 875 kg/m^3 in the current study.

2.2. Acceleration of TEHL Model

While the pressure and film thickness initial approximations in the TEHL simulation framework are somewhat close to the final desired solutions in isothermal cases, there are no empirical approaches to approximate the temperature profile in the three-dimensional space. This means that very often the initial temperature profile, and by extension the pressure and film thickness profiles, in TEHL simulations are not close to the final solutions. Hence, the simulation requires many iterations to converge, leading to long simulation times.

In this study, trained ML models are used to predict values of rigid separation (h_0), maximum temperature (T_{max}), mid-film friction coefficient (μ), pressure (p) profiles, liquid film fraction (θ) profiles and fluid temperature (T) profiles. The predicted values of h_0, p, θ and T are used as initial approximations for the TEHL simulations.

Three simulation frameworks will be compared: 1) classical FVM framework (from now on referred to as 'Classical FW'), 2) hybrid NN-FVM framework with initial fluid temperature prediction ('Hybrid FW 1') and 3) hybrid NN-FVM framework with initial rigid separation, pressure, liquid film fraction and fluid temperature prediction ('Hybrid FW 2'). To quantify the amount of acceleration of each simulation, an easily interpretable acceleration factor metric is used as defined below:

$$\text{Acceleration factor} = \frac{t_{classical}}{t_{hybrid}}, \quad (15)$$

where $t_{classical}$ indicates the simulation time of the classical TEHL framework and t_{hybrid} indicates the simulation time of the hybrid NN-TEHL frameworks. Values above 1 indicate that the hybrid NN-FVM framework simulation time is smaller than the classical FVM framework simulation time, while values below 1 indicate the opposite. For example, an acceleration factor value of 2 denotes that the hybrid NN-FVM framework is twice as fast as the classical FVM framework.

The simulation time is defined as the wall-clock time elapsed during the execution of each simulation run, measured using MathWorks MATLAB's *tic* and *toc* functions, which respectively start and stop a stopwatch timer. The *tic* function is called after all input parameters have been defined. The *toc* function is called immediately when global convergence has been reached and excludes any time spent on any post-processing steps like file-saving. All TEHL simulations were performed using MathWorks MATLAB R2024a on a computer equipped with a 13th Gen Intel® Core™ i7-13850HX (base speed 2.10 GHz), 64 GB of RAM and running the Windows 11 Enterprise operating system. A uniform $129 \times 129 \times 11$ grid is used in both fluid and solid domains in all simulations.

2.3. Training and Testing Data Generation

The ML models in this study use 12 parameters as inputs. These include operating conditions like applied load, entrainment speed, slide-to-roll ratio (SRR) and ambient temperature. They also include properties of the solid bodies like effective radius of curvature, Young's modulus, Poisson's ratio, thermal conductivity, specific heat capacity and density as well as properties of the lubricant like reference viscosity and pressure-viscosity coefficient.

For each input parameter a range of values is chosen to reflect the range of operating conditions and materials used in a mini traction machine (MTM) [32]. Table 1 shows the range of values used for each input parameter. Similarly to [22], the Latin Hypercube Sampling (LHS) method was used to sample data points with the desired range of values using the MATLAB function *lhsdesign*.

During the generation of the samples, some of them are discarded to ensure fully flooded EHL conditions as well as to ensure that the maximum Hertzian pressures do not exceed the pressures in the MTM. The expected minimum film thickness for each sample is calculated using Hamrock and Dowson minimum film thickness formula for point contacts given below [28]:

$$h_{min,HD} = 3.63 R U^{*0.68} G^{*0.49} W^{*-0.073}(1 - e^{-0.68}). \quad (16)$$

The maximum Hertzian pressure was calculated using the equation below [33]:

$$p_{Hertz} = \left(\frac{3WE'^2}{2R^2\pi^3}\right)^{1/3}. \quad (17)$$

In total, 600 samples were generated using the method above, with 85% of them (510 samples) used for training the ML model and 15% (90 samples) used for testing its performance. The inputs that are constant across all simulations are summarised in Table 2.

The dimensionless load (M) and material (L) parameters were used to group the generated samples. The expressions for M and L for point contacts are given below [34, 35]:

$$\begin{aligned} M &= W^*(2U^*)^{-3/4}, \\ L &= G^*(2U^*)^{1/4}. \end{aligned} \quad (18)$$

2.4. Machine Learning Models

Similarly to the studies of Marian et al. [20] and Tošić et al. [22], this study focuses on ANNs which consist of layers of interconnected neurons. ANNs are used because they exhibit superior prediction accuracy compared to other algorithms like linear regression, regression trees, support vector machines (SVMs), ensembles of trees and Gaussian Process Regression

Table 1: Training and testing sample parameters

Parameter type	Parameter	Min.	Max.
Operating conditions	Applied load, W [N]	10	100
	Entrainment speed, u_{ent} [m/s]	0.01	3.00
	Slide-to-roll ratio, SRR [-]	0	2
	Ambient temperature, T_{amb} [°C]	20	40
Solid properties	Effective radius of curvature, R [m]	0.003	0.020
	Young's modulus of solid bodies, $E_{1,2}$ [GPa]	70	400
	Poisson's ratio of solid bodies, $\nu_{1,2}$ [-]	0.15	0.35
	Thermal conductivity of solid bodies, $k_{1,2}$ [W/(m K)]	1	250
	Specific heat capacity of solid bodies, $c_{p1,2}$ [J/(kg K)]	500	900
	Density of solid bodies, $\rho_{1,2}$ [kg/m^3]	1500	8000
Lubricant properties	Lubricant reference viscosity, η_0 [Pa s]	0.01	0.10
	Lubricant pressure-viscosity coefficient, α_p [1/Pa]	10×10^{-9}	25×10^{-9}

Table 2: Constant parameters

Parameter type	Parameter	Value
Lubricant properties	Fluid specific heat capacity, c_p [J/(kg K)]	2000
	Fluid thermal conductivity, k [W/(m K)]	0.14
	Fluid coefficient of thermal expansion, β [1/K]	6.5×10^{-4}
	Eyring shear stress, τ_E [Pa]	1×10^7
	Reference density, ρ_0 [kg/m^3]	875
Simulation parameters	Mesh size for fluid and solid domains, $N_x \times N_y \times N_z$	$129 \times 129 \times 11$
	Computational domain	$-3.6 \leq x/a \leq 1.52$, $-1.6 \leq y/a \leq 1.6$ *
	Pressure-liquid film fraction convergence criterion, $e^{p,\theta}$	1×10^{-5}
	Load convergence criterion, e^W	1×10^{-4}
	Temperature convergence criterion, e^T	1×10^{-5}
	Global pressure convergence criterion, e^p_{global}	1×10^{-3}
	Global temperature convergence criterion, e^T_{global}	1×10^{-3}

*$a = [3WR/(2E')]^{1/3}$ is the Hertzian contact radius.

(GPR) [22]. Figure 3 shows an example of an ANN with 12 inputs, one hidden layer, six neurons in the hidden layer and one output.

According to Guresen and Kayakutlu [36], each neuron consists of three basic components: a synaptic weight which determines the influence of the input to the neuron, a summing function that combines each input to the neuron with the corresponding weight of that input and an activation function to produce an output. During training, the ANN automatically adjusts the synaptic weights in each neuron as well as a bias term which is also included in the summing function and is used to adjust the output of each neuron independently of the inputs. The training continues until the residual, or loss, between the predicted and desired outputs is minimised.

In this study, six different ANNs are trained in MathWorks MATLAB R2024b to predict the values of h_0, T_{max}, μ and the profiles of p, θ and T. All ANNs receive the same 12 parameters as inputs.

2.4.1. Standardisation of Variables

Before they are used for training, all input variables are standardised using the mean and standard deviation of the whole input dataset. The standardised variables are calculated in the following way:

$$\tilde{y}_i = \frac{y_i - \overline{y}}{\sigma}, \tag{19}$$

where $\tilde{y}_i$ is the standardised variable, y_i is the non-standardised variable, $\overline{y}$ is the mean of the whole dataset and σ is the standard deviation of the whole dataset.

Output values for h_0, T_{max}, μ, p and θ are standardised in the same way as the input variables. In addition, the values of p are normalised by dividing them by the maximum Hertzian pressure (p_{Hertz}). Output values for T are scaled using the minimum and maximum values of *each* sample as shown below:

$$\tilde{T}_i = \frac{T_i - T_{amb}}{T_{max} - T_{amb}}, \tag{20}$$

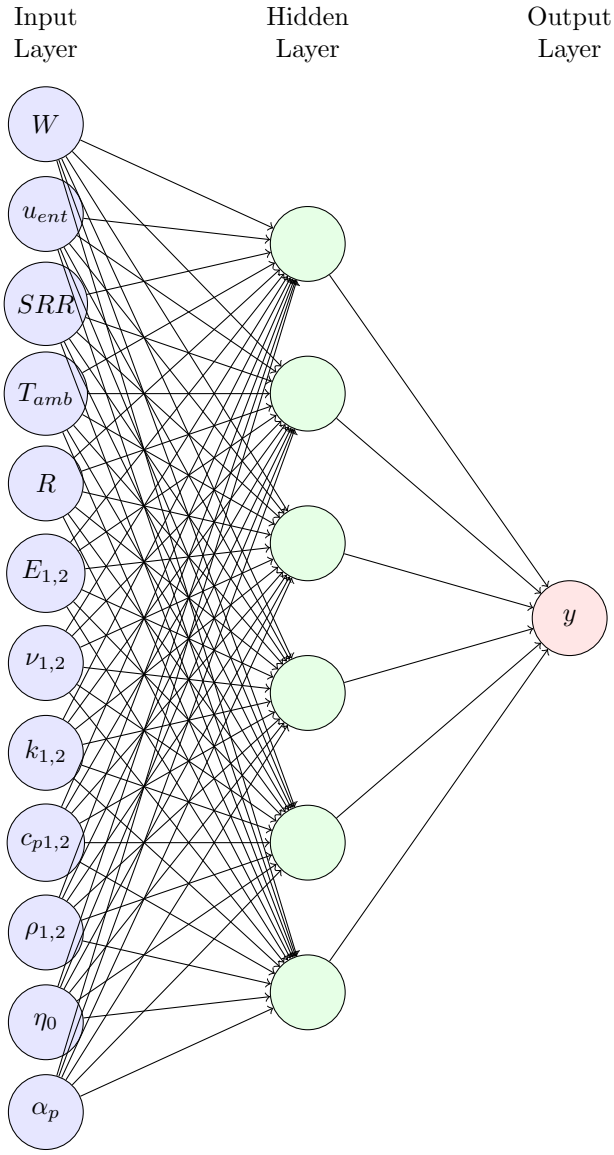


Figure 3: Artificial Neural Network schematic

where $\tilde{T}_i$ is the rescaled temperature value, T_i is the original temperature value, T_{amb} is the ambient temperature of each sample and T_{max} is the maximum temperature of each sample.

In the testing phase, when predicting the above variables using inputs that were not used for training, the outputs values for h_0, T_{max}, μ, p and θ can be destandardised using the corresponding mean and standard deviation used during training, while destandardisation of the T values requires use of the predicted T_{max} values.

2.4.2. Neural Network Architecture

Each hidden layer is constructed in MATLAB using a *fullyConnectedLayer*, a *swishLayer* and a *dropoutLayer* with a given feature dropout probability to avoid overfitting. In this study a Swish activation function is used instead of more traditional choices like the Rectified Linear Unit (ReLU) or the Hyperbolic Tangent (tanh) activation functions. According to Szandała [37], the Swish activation function can perform better compared to the widely used ReLU function in deep networks. The ReLU function can suffer from dead neuron problems due to zero gradients for negative inputs, while the Swish function maintains a small but non-zero gradient across the domain. Compared to the tanh function, which can lead to vanishing gradients due to saturation for large inputs, the Swish function does not suffer from this issue. The only disadvantage of the Swish function is its higher computational cost compared to the ReLU function for both feedforward and backpropagation problems [37].

To handle the inputs, a *featureInputLayer* is used with a feature size equal to the number of input parameters (12). A final *fullyConnectedLayer* with size equal to the number of output values of each ANN is used in addition to a *regressionLayer* to handle the outputs. The adaptive moment estimation (ADAM) algorithm is used for training the ANNs because it is computationally efficient, has small memory requirements and is suited to problems with large numbers of data or parameters [38].

The metric used to quantify the prediction accuracy of the ANNs is the coefficient of determination (R^2), which is calculated using the equation below:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \tag{21}$$

where y_i is the actual value of a variable, $\hat{y}_i$ is the predicted ANN value and N is the number of output values of each ANN for each sample. For example, for the rigid separation ANN, $N = 1$ since a scalar value is predicted. For the pressure profile ANN, $N = N_x \times N_y = 129 \times 129 = 16,641$. For the temperature profile ANN, $N = N_x \times N_y \times N_z = 129 \times 129 \times 11 = 183,051$.

Since the coefficient of determination (R^2) is unbounded on the negative side, negative values can be misleading, potentially suggesting an inverse relationship rather than simply a poor prediction. In this study, such inverse relationships are not expected and a negative R^2 would instead indicate a highly inaccurate prediction. To address this, a rescaled R^2 metric (R^2_{rs}),

bounded between 0 and 1, is used. This is calculated using the following equation:

$$R^2_{rs} = \frac{1}{2 - R^2}, \tag{22}$$

An R^2_{rs} of 1 corresponds to an R^2 of 1 and signifies a perfect correlation between the predicted and true values. An R^2_{rs} of 0.5 corresponds to an R^2 of 0, and indicates that the model is no better than just using the mean of the dataset to make a prediction. Values of R^2_{rs} below 0.5 correspond to negative R^2 values and indicate that the model performs worse than just using the mean of the whole dataset to make a prediction.

2.4.3. Hyperparameter Optimisation

Bayesian optimisation is used to find a set of ANN hyperparameters that results to the best performing ANN. This is the ANN that results to the highest global R^2_{rs} for a specific dataset. In the case of two- or three-dimensional predicted profiles (p, θ, T), this is found by computing the mean R^2_{rs} of all samples in the dataset, and in the case of scalar predicted values (h_0, T_{max}, μ) this is just the R^2_{rs} of the dataset. A Bayesian optimisation algorithm (MATLAB's *bayesopt* in this study) performs many iterations of the ANN training with different hyperparameters, each time observing which areas of the hyperparameter space lead to the best performing ANN in terms of global R^2_{rs}. Instead of randomly choosing different hyperparameters, it learns from the previous iterations and does not waste time on hyperparameters that lead to suboptimal performance [39]. Table 3 summarises the hyperparameter search space used in the Bayesian optimisation process.

Table 3: Hyperparameter search space

Parameter	Range
Number of hidden layers	1-5
Number of neurons per layer	12-512
Mini batch size	2-64
Dropout layer probability	0.01-0.5

3. Results and discussion

3.1. Training and Testing Data

As mentioned previously, during the generation of the training and testing samples, some samples were discarded to ensure fully-flooded conditions as well as to ensure that the maximum Hertzian pressures do not exceed the pressures in the MTM. More specifically, samples with a theoretical minimum film thickness of less than 30 nm (from Eq. (16)) and a maximum Hertzian pressure greater than 1.5 GPa (from Eq. (17)) were discarded. Moreover, as was the case in the study of Tošić et al. [22], samples with a maximum Hertzian pressure smaller than 0.05 GPa were discarded to ensure elastohydrodynamic and not hydrodynamic lubrication conditions. Lastly, samples with a maximum Hertzian pressure greater than 0.6 GPa and Young's modulus smaller than 100 GPa were discarded to avoid plastic deformation of soft materials. New samples were generated in place of the discarded samples so that 600 samples in total were tested.

Figures 4 and 5 show the rigid separation values of the training and testing samples respectively across the parameter space normalised by the approximation for the rigid separation calculated using the Hamrock and Dowson central film thickness formula (Eq. (1)). As seen in the figures, the actual rigid separation of the two bodies is only close to the Hamrock and Dowson initial approximation at low values of M and high values of L.

The values of the Moes dimensionless parameters for the training dataset range from 5.83 to 1599.40 in the case of M and from 2.06 to 27.22 in the case of L. For the testing dataset, the M values range from 4.16 to 1355.50 and the L values range from 1.64 to 28.09. This range is large enough to cover both the piezoviscous-rigid regime and the piezoviscous-elastic regime as classified by Moes [40] and Nijenbanning et al. [41].

Figures 6 and 7 show the normalised maximum hydrodynamic pressure values of the training and testing samples respectively across the parameter space. These figures show that the maximum pressure in many samples is significantly higher than the maximum pressure of the Hertzian profile that is used as an initial approximation in the classical TEHL simulations. The high dimensionless maximum pressures occurring at low M and high L values are the result of the pressure spike at the outlet region of the contact which cannot be predicted using Hertzian theory.

Similarly, Figures 8 and 9 show the normalised maximum temperature values of the training and testing samples respectively across the parameter space. For many samples, the normalised maximum temperature is a lot higher than the ambient temperature used to initialise the classical TEHL simulations, especially as L increases.

Finally, Figures 10 and 11 show the mid-film friction coefficient values of the training and testing samples respectively across the parameter space. Generally, the friction coefficient values increase as L increases.

It is evident in the above results that the initial approximations for rigid separation, pressure and temperature in classical TEHL simulations can be greatly improved upon, potentially allowing for shorter simulation times.

3.2. Optimised Neural Network Architecture

Table 4 summarises the results of the Bayesian optimisation for each ANN. The hyperparameters in this table are used to train the respective ANNs.

3.3. Accuracy of Neural Network Predictions

Figure 12a compares the rigid separation ANN predictions with the rigid separation values obtained from the FVM simulations for the testing samples dataset. The figure shows that the NN predictions are very close to the values obtained from the FVM simulations with an R^2_{rs} value of 0.997 for the whole testing dataset.

Figure 12b compares the maximum temperature ANN predictions with the maximum temperature values obtained from the FVM simulations for the testing samples dataset. Once

Table 4: Optimal hyperparameters for each neural network.

Parameter	h_0 ANN	T_{max} ANN	μ ANN	p ANN	θ ANN	T ANN
Number of hidden layers	4	4	4	5	2	2
Number of neurons in each layer	249, 108, 41, 190	509, 419, 258, 20	204, 252, 254, 121	264, 12, 391, 282, 377	225, 25	466, 135
Mini batch size	30	33	51	64	11	55
Dropout layer probability	0.0194	0.0201	0.0407	0.0109	0.0111	0.4889

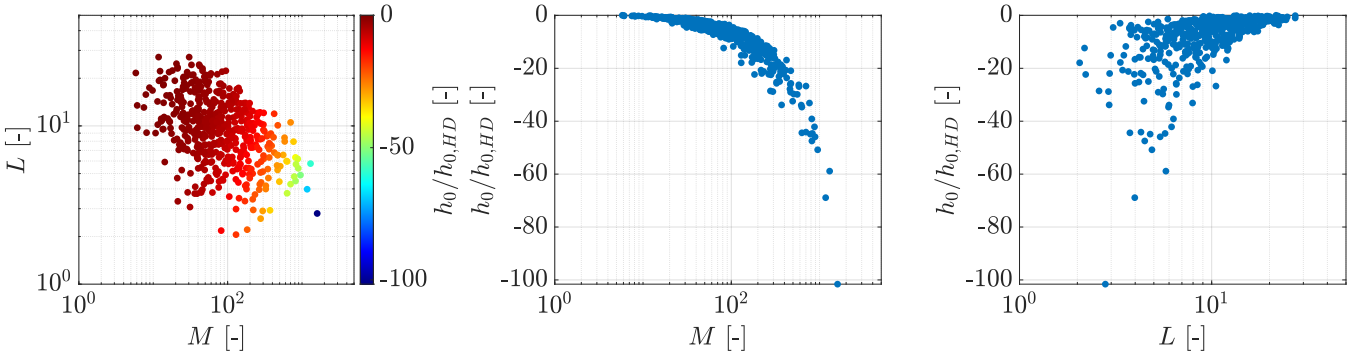


Figure 4: Normalised rigid separation ($h_0/h_{0,HD}$) of training samples across parameter space.

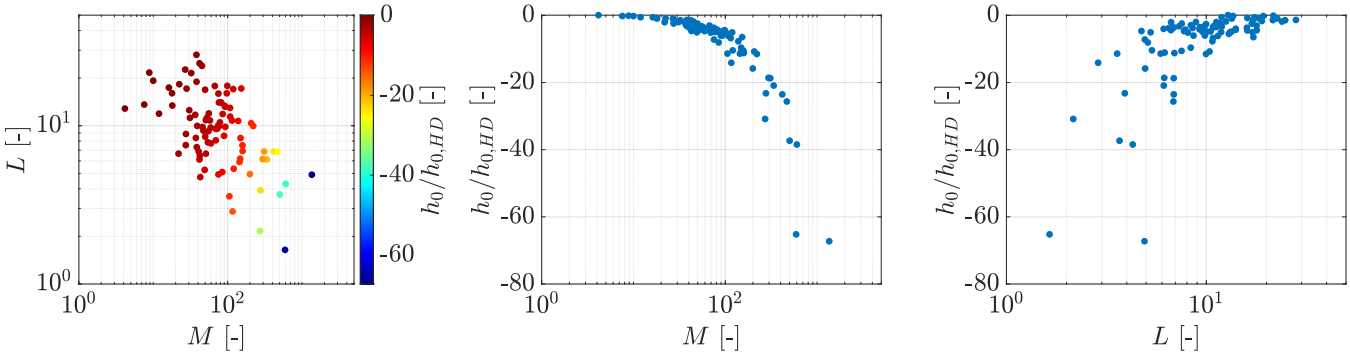


Figure 5: Normalised rigid separation ($h_0/h_{0,HD}$) of testing samples across parameter space.

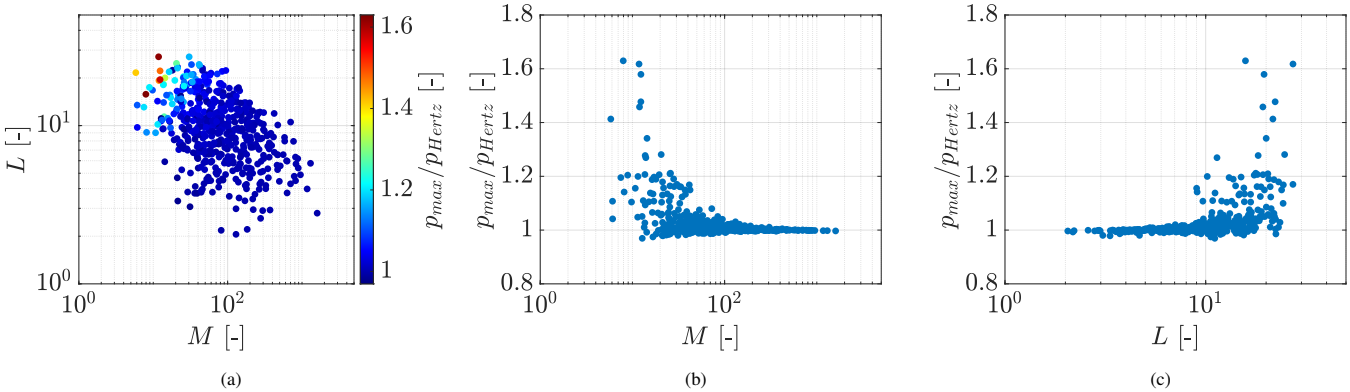
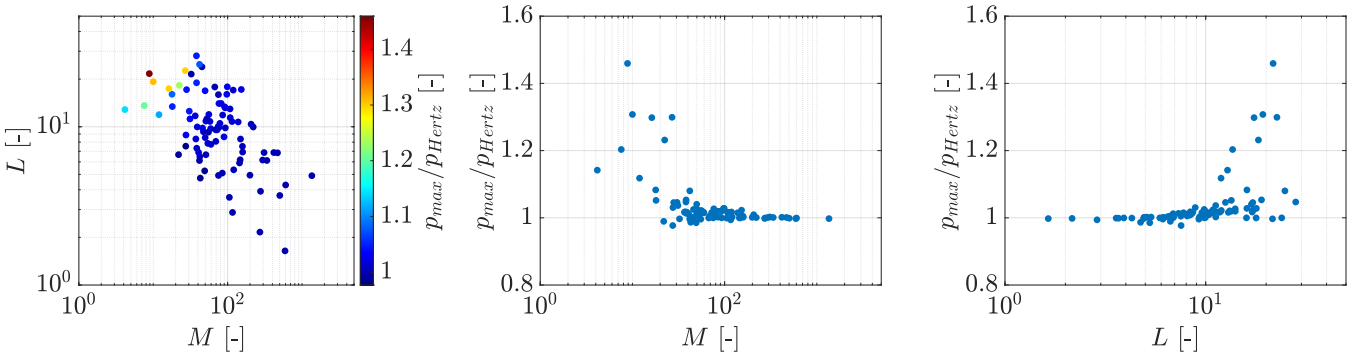
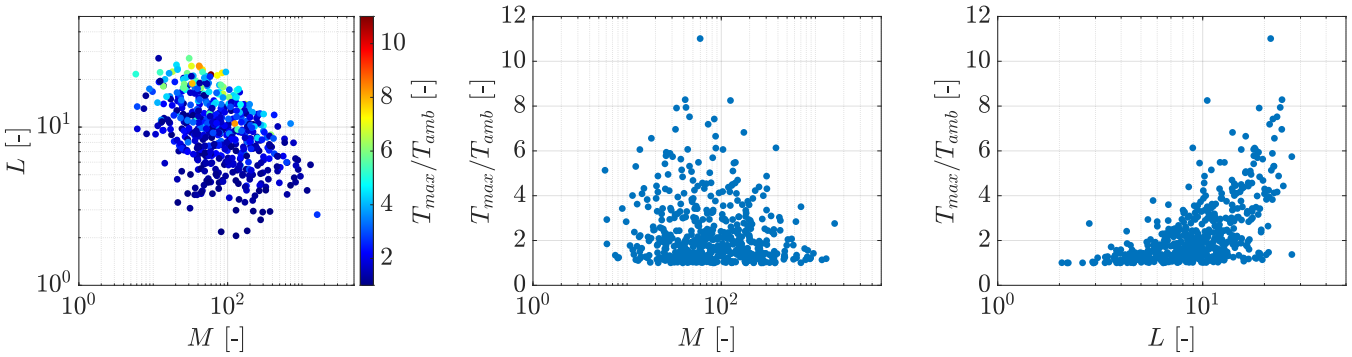


Figure 6: Normalised maximum hydrodynamic pressure (p_{max}/p_{Hertz}) of training samples across parameter space.



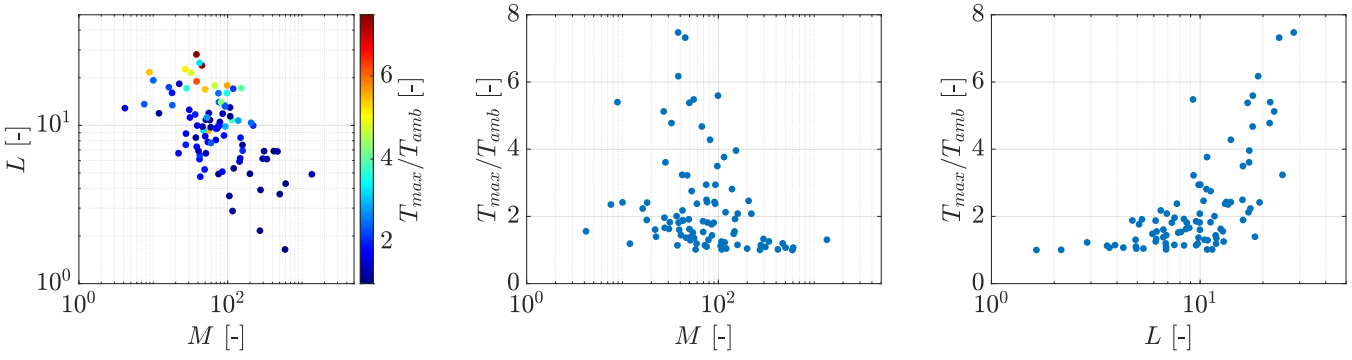
(a) (b) (c)

Figure 7: Normalised maximum hydrodynamic pressure (p_{max}/p_{Hertz}) of testing samples across parameter space.



(a) (b) (c)

Figure 8: Normalised maximum temperature (T_{max}/T_{amb}) of training samples across parameter space.



(a) (b) (c)

Figure 9: Normalised maximum temperature (T_{max}/T_{amb}) of testing samples across parameter space.

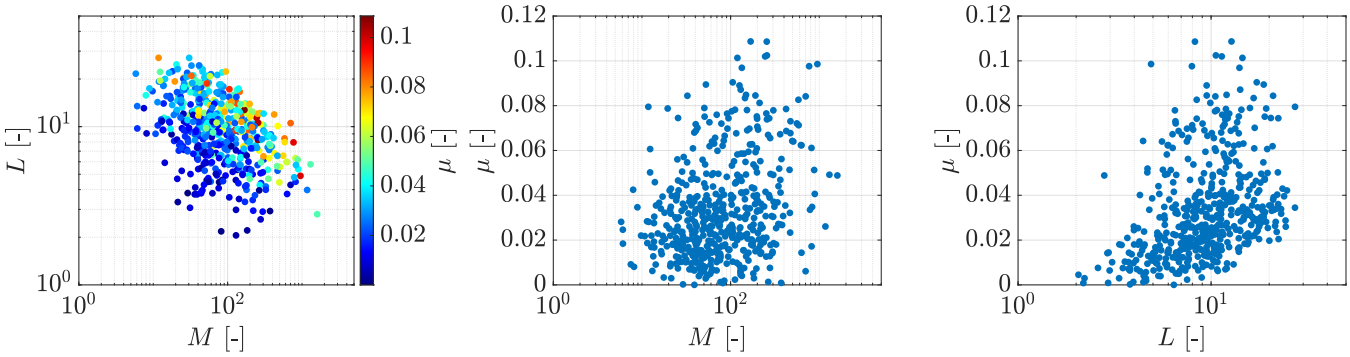


Figure 10: Mid-film friction coefficient (μ) of training samples across parameter space.

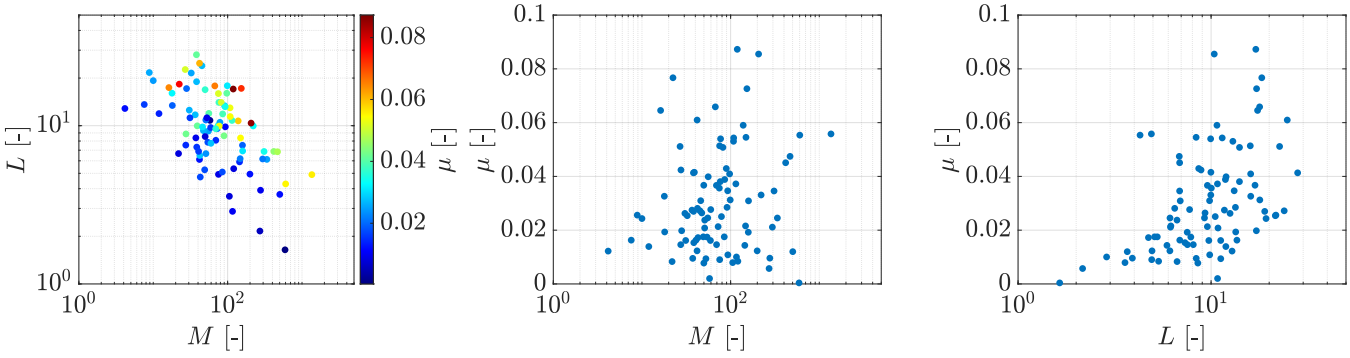
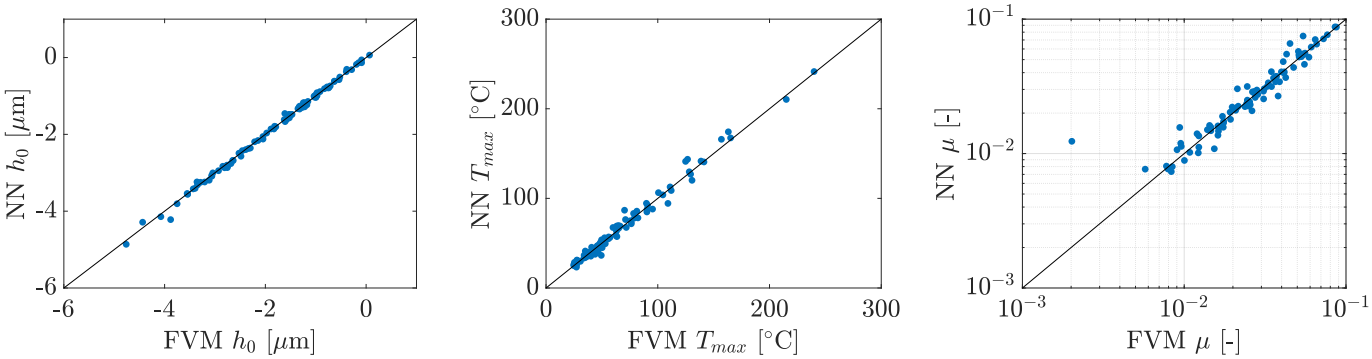


Figure 11: Mid-film friction coefficient (μ) of testing samples across parameter space.



(a) Rigid separation (R^2_{rs} = 0.997)

(b) Maximum temperature (R^2_{rs} = 0.985)

(c) Mid-film friction coefficient (R^2_{rs} = 0.939)

Figure 12: Accuracy of (a) rigid separation (h_0), (b) maximum temperature (T_{max}) and (c) mid-film friction coefficient (μ) predictions.

again, the NN predictions are very close to the values obtained from the FVM simulations with an R^2_{rs} value of 0.985 for the whole testing dataset.

With regard to the prediction of the mid-film friction coefficient, Figure 12c shows that all but one NN predictions are in very good agreement with the values obtained from the FVM simulations with an R^2_{rs} value of 0.939 for the whole testing dataset. The one exception is the prediction corresponding to the lowest friction coefficient, which also corresponds to the sample with the lowest L value, as shown in Figure 11c. Notably, this sample has an L value of 1.64, which lies outside the training range of 2.06 to 27.22. This likely contributes to the reduced prediction accuracy. Additionally, it is possible that the network tends to prioritize minimising errors for larger friction coefficients, which dominate the dataset in terms of magnitude, making it less sensitive to accurately capturing smaller values.

Figure 13 shows the values of R^2_{rs} for the pressure profile ANN prediction of each testing sample across the parameter space. All NN predictions are very accurate with a mean R^2_{rs} for the whole dataset of 0.997. The predictions seem to be less accurate, relatively speaking, for low M and high L values, possibly due to the presence of a prominent outlet pressure spike for those samples as discussed earlier. These localised nonlinearities are naturally more difficult for the ANNs to capture.

Figures 14 and 15 show the most and least accurately predicted pressure profiles respectively. In the best-case scenario, the ANN prediction closely follows the FVM solution across the entire domain, and the difference profile reveals only minor deviations up to about 0.03 units in magnitude. In the worst-case scenario, while the overall trend of the pressure distribution is still captured, the largest discrepancies appear near the outlet region, where the pressure spike occurs with maximum deviations of up to about 0.4 units, 10 times greater than the deviations in the best-case scenario. This confirms that the ANN is less accurate in its attempts to capture sharp gradients or localised features.

The predictions for liquid film fraction seem to follow a similar trend, as shown in Figure 16, with a mean R^2_{rs} for the whole dataset of 0.997. The high prediction accuracy can be partly attributed to the fact that both the normalised pressure and liquid film fraction profiles are naturally bounded between 0 and 1, which makes it easier for the ANN to learn the spatial patterns without ignoring profiles with smaller absolute magnitudes.

Figures 17 and 18 show the most and least accurately predicted liquid film fraction profiles respectively. In general, the ANN captures the spatial variations well, although some localised noise is observed, particularly in regions where $\theta = 1$. These regions tend to have very low variation, resulting in flat profiles. These regions, characterised by small gradients in the datasets, are given less emphasis during training since the network prioritises regions where there is more variation in the data. This may lead to reduced accuracy and the appearance of localised noise.

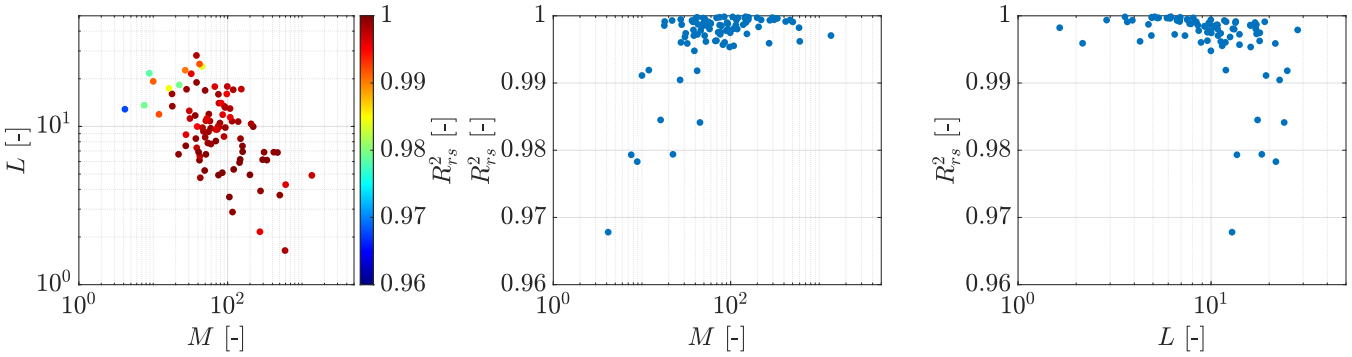
In contrast to the pressure and liquid film fraction profiles, predicting the three-dimensional temperature profiles using ANNs seems to be more challenging. Figure 19 shows the values of R^2_{rs} for the temperature profile ANN prediction of each testing sample across the parameter space. The mean R^2_{rs} for the whole dataset is 0.873, which is lower than that of the pressure or liquid film fraction predictions. The reduced accuracy is likely due to the fact that there is no empirical quantity (as is the case with pressure) that could be used to normalise the temperature profiles so that they all lie in a similar range. Instead, the temperature profiles are normalised by the *known* maximum temperature of each sample during training and then denormalised using the *predicted* maximum temperature during testing. The predicted maximum temperature has its own associated error, which, despite it being small as already discussed for samples with small temperature variations, could lead to relatively large deviations between true and predicted profiles, hence extremely small R^2_{rs} values.

Figures 20 and 21 show the most and least accurately predicted temperature profiles respectively at the middle of the film ($z/h = 0.5$) and Figures 22 and 23 show the corresponding temperature variations across the z-direction at $y = 0$. In the best-case temperature prediction, the ANN closely matches the FVM solution, with a maximum deviation of about 0.7°C, which is a relatively small error (1.29%). This indicates that the model performs well when the temperature profile has clear and noticeable variations, accurately capturing the temperature distribution in the three-dimensional space. For the worst-case temperature prediction, the corresponding FVM solution exhibits very low variation in temperature. In contrast, the ANN prediction shows a temperature *reduction* instead of predicting a nearly flat profile. This discrepancy can be attributed to the fact that the ANN likely prioritised learning from temperature profiles with greater variation (or greater maximum temperature) during the training phase. The network may have learned to focus on these more varied patterns and failed to generalise well to the flat profiles. This hypothesis is confirmed by Figure 24, which clearly shows that the least accurate predictions are the ones made for samples with very small temperature variations (low maximum temperatures). This again highlights a limitation in the ANN's ability to handle profiles with small variations.

In summary, ANNs can be used to predict various parameters in tribological systems with high accuracy. These parameters can range from scalar values, such as rigid separation, maximum temperature, and friction coefficient, to two-dimensional profiles like pressure and liquid film fraction distributions, or even three-dimensional profiles like temperature distributions. The key for accurate prediction of two- and three-dimensional profiles appears to lie in appropriate normalisation, ensuring that the desired outputs have similar magnitudes. By normalising the data effectively, the network can prevent data samples with higher output magnitudes from dominating the training process. This is particularly important for predicting temperature profiles, where issues such as small temperature variation (low maximum temperatures) can lead to inaccurate predictions.

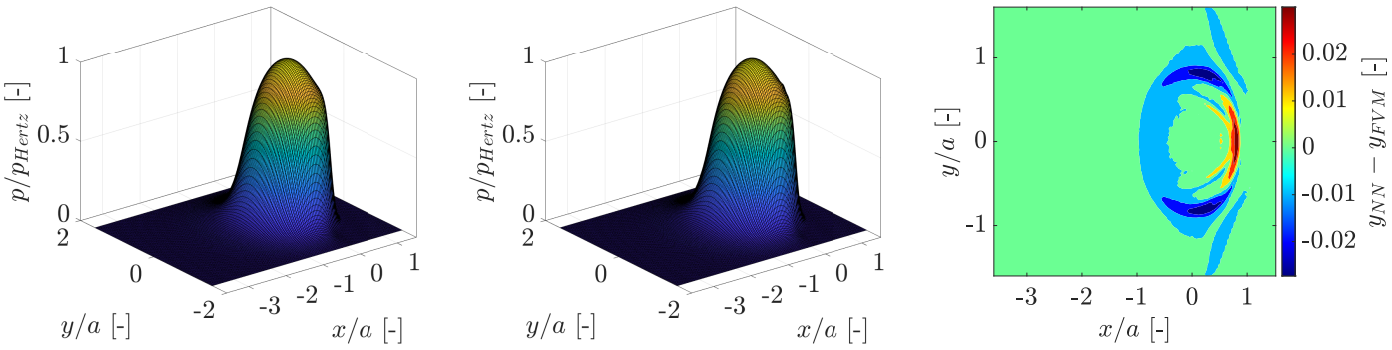
3.4. Simulation Time Comparison

Figure 25 shows the Classical FW simulation times across the parameter space. The mean simulation time of the whole



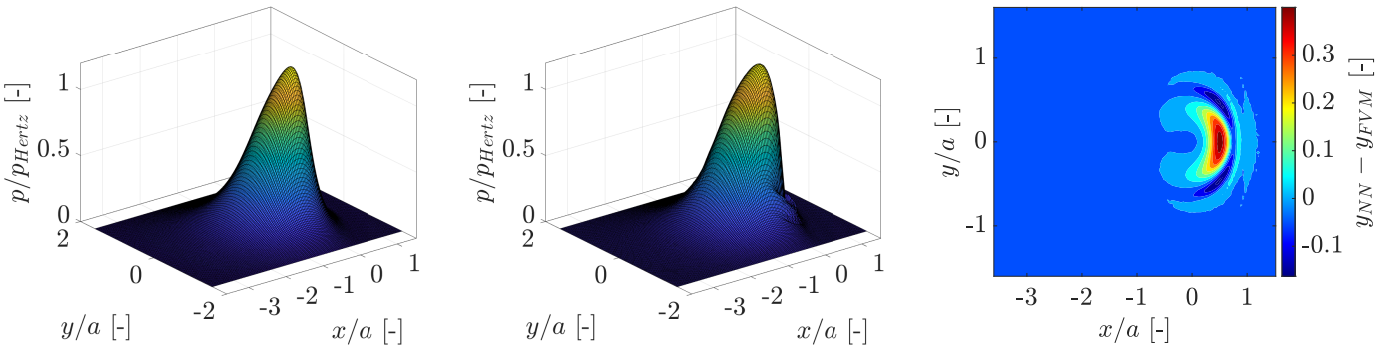
(a) (b) (c)

Figure 13: Accuracy of pressure (p) profile predictions across parameter space (mean R^2_{rs} = 0.997).



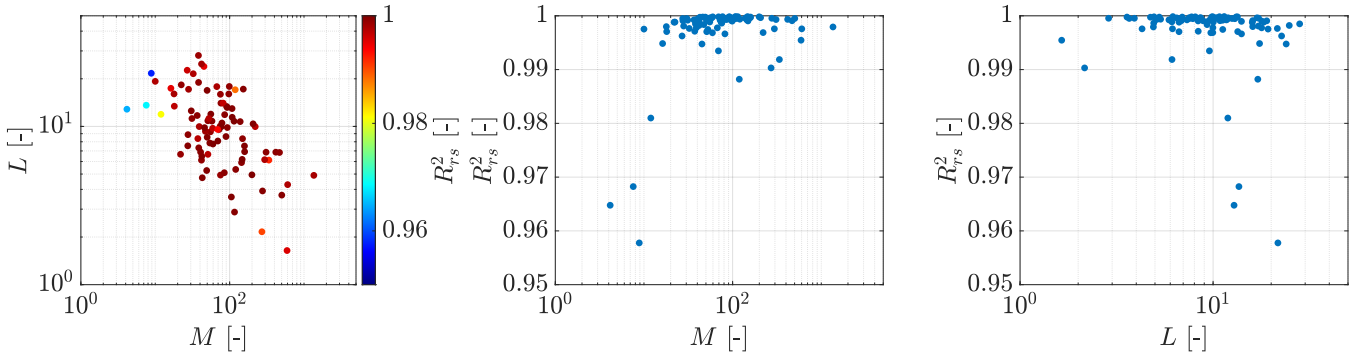
(a) FVM solution (b) ANN prediction (c) Difference between NN and FVM profiles

Figure 14: Example of most accurately predicted pressure profile (Testing sample 25, R^2_{rs} = 1.000).



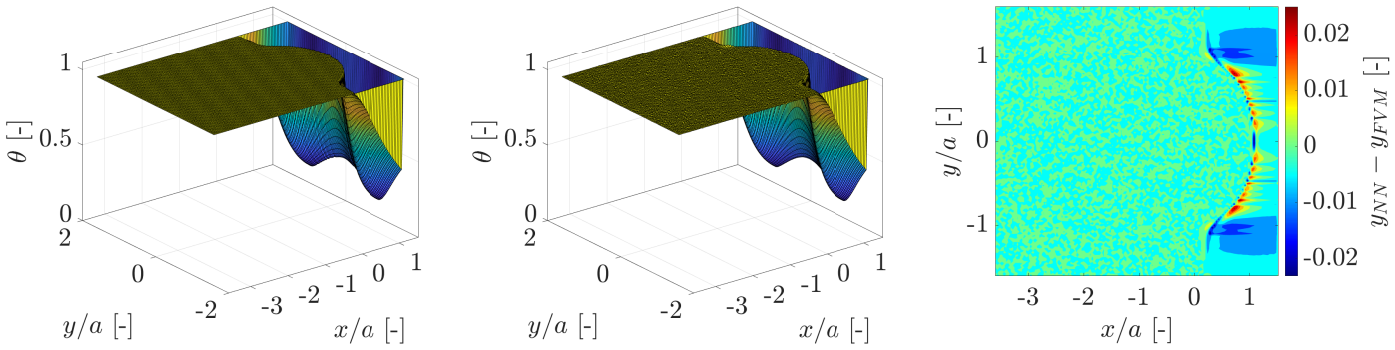
(a) FVM solution (b) ANN prediction (c) Difference between NN and FVM profiles

Figure 15: Example of least accurately predicted pressure profile (Testing sample 11, R^2_{rs} = 0.967).



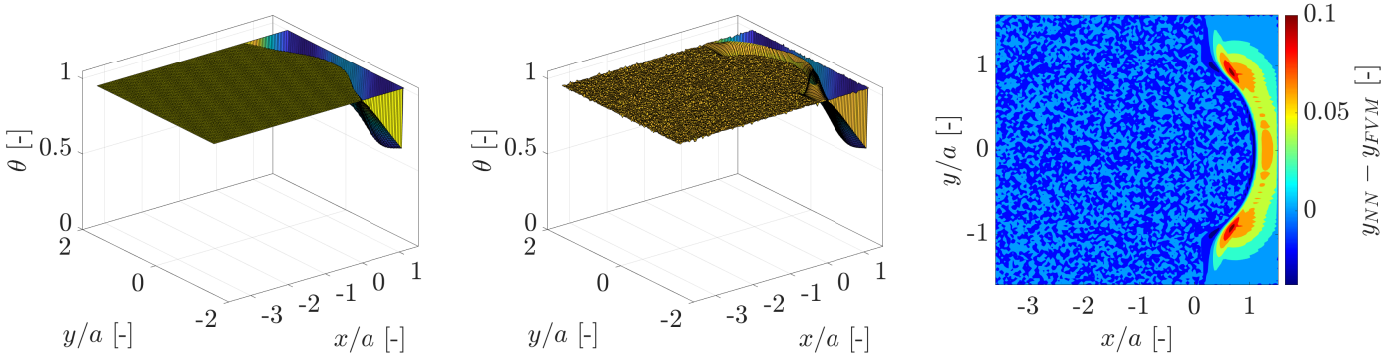
(a) (b) (c)

Figure 16: Accuracy of liquid film fraction (θ) profile predictions across parameter space (mean R^2_{rs} = 0.997).



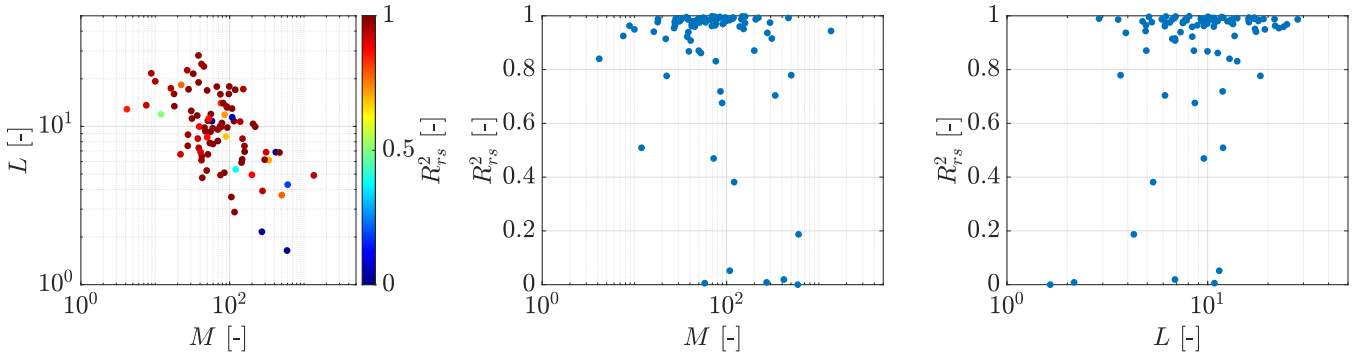
(a) FVM solution (b) NN prediction (c) Difference between NN and FVM profiles

Figure 17: Example of most accurately predicted liquid film fraction profile (Testing sample 10, R^2_{rs} = 1.000).



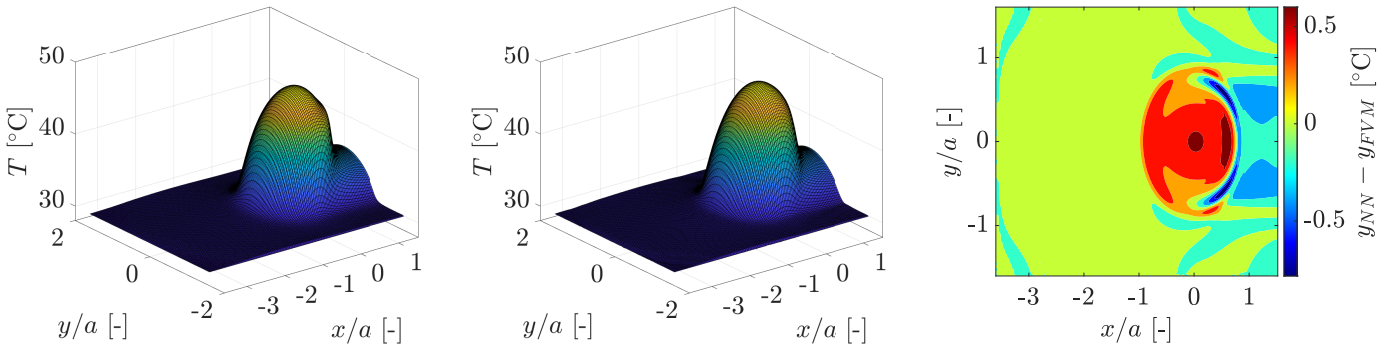
(a) FVM solution (b) NN prediction (c) Difference between NN and FVM profiles

Figure 18: Example of least accurately predicted liquid film fraction profile (Testing sample 5, R^2_{rs} = 0.956).



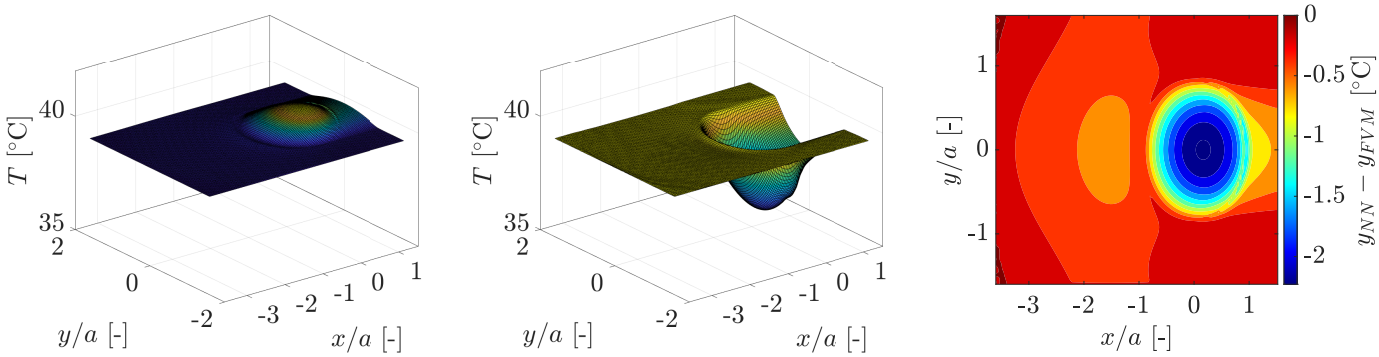
(a) (b) (c)

Figure 19: Accuracy of temperature (T) profile predictions across parameter space (mean R^2_{rs} = 0.873).



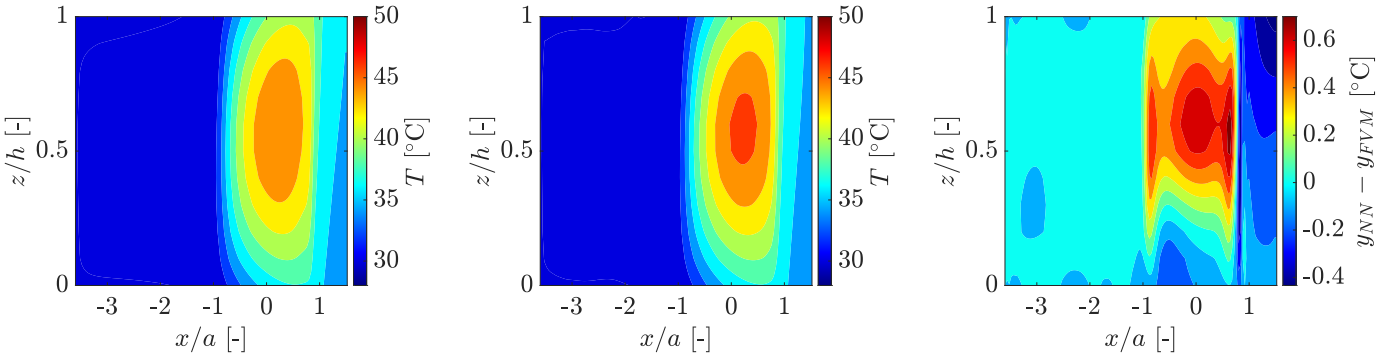
(a) FVM solution (b) NN prediction (c) Difference between NN and FVM profiles

Figure 20: Example of most accurately predicted mid-film temperature profile (z/h = 0.5) (Testing sample 10, R^2_{rs} = 0.998).



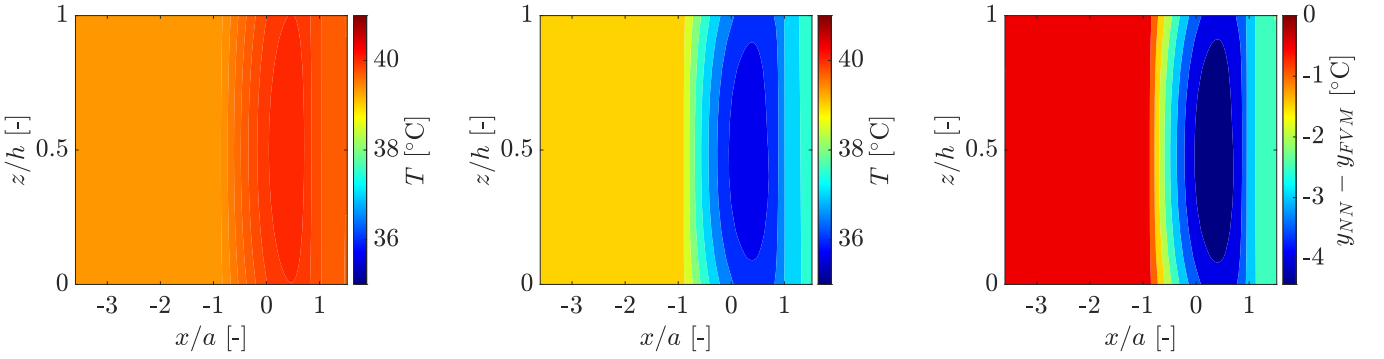
(a) FVM solution (b) NN prediction (c) Difference between NN and FVM profiles

Figure 21: Example of least accurately predicted mid-film temperature profile (z/h = 0.5) (Testing sample 47, R^2_{rs} = 0.019).



(a) FVM solution (b) NN prediction (c) Difference between NN and FVM profiles

Figure 22: Example of most accurately predicted temperature profile across the z-direction ($y = 0$) (Testing sample 10, $R^2_{rs} = 0.998$).



(a) FVM solution (b) NN prediction (c) Difference between NN and FVM profiles

Figure 23: Example of least accurately predicted temperature profile across the z-direction ($y = 0$) (Testing sample 47, $R^2_{rs} = 0.019$).

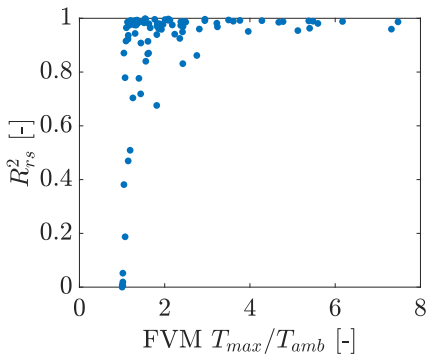


Figure 24: Effect of normalised maximum temperature on predicted temperature profile accuracy.

testing dataset using this framework is 19.43 minutes. Comparing with Figures 8 and 9, which show the maximum temperatures across the parameter space, it is evident that the classical FW simulations take more time to converge at high L values where the maximum temperatures are higher. This is expected as the higher the final temperature is for a given simulation, the further away the final solution is from the initial ambient temperature assumption, hence more iterations are required for the solution to converge.

Figure 26 shows the Hybrid FW 1 simulation times across the parameter space. The mean simulation time of the whole testing dataset using this framework is 14.42 minutes which is 25.78% less than the mean simulation time of the Classical FW simulations. The Hybrid FW 1 simulation time results show that predicting the initial temperature profile using ANNs can have a significant acceleration effect on the TEHL simulations. Figure 27 shows the acceleration factor values of the Hybrid FW 1 simulations across the parameter space, with all but six simulations showing some amount of acceleration, with some Hybrid FW 1 simulations being up to even twice as fast as the Classical FW simulations.

Figure 28 shows the Hybrid FW 2 simulation times across the parameter space. The mean simulation time of the whole testing dataset using this framework is 12.09 minutes which is 37.78% less than the mean simulation time of the Classical FW simulations. The Hybrid FW 2 simulation results show that predicting the initial rigid separation value, pressure profile and liquid film fraction profile using ANNs, in addition to predicting the initial temperature profile, can amplify the acceleration effect. In Figure 29 it is seen that the acceleration factor values are even higher, with some Hybrid FW 2 simulations being up to three times as fast as the Classical FW simulations. Once again, six Hybrid FW 2 simulations actually take longer than the corresponding Classical FW simulations.

As seen in Figures 30 and 31, the six Hybrid FW 1 and Hybrid FW 2 simulations that take longer than the Classical FW simulations are the simulations with the lowest normalised maximum temperature, hence the smallest temperature variations. As mentioned previously, these correspond to samples whose temperature profiles are not very accurately predicted by the ANNs. Figures 32 and 33 show the effect that temperature prediction accuracy (in terms of R^2_{rs}) has on the acceleration factors of the Hybrid FW 1 and Hybrid FW 2 simulations respectively. In the case of the Hybrid FW 1 simulations, five of the six simulations that take longer than the Classical FW simulations have R^2_{rs} values for the temperature profile prediction smaller than 0.2. In the case of the Hybrid FW 2 simulations, all six simulations that take longer than the Classical FW simulations have R^2_{rs} values for the temperature profile prediction smaller than 0.4. Conversely, the simulations showing the highest acceleration factors are the ones with a temperature profile prediction accuracy close to 1. Clearly, simulation acceleration is linked to the ANN prediction accuracy.

While the present study demonstrates the accuracy of ANNs in predicting various tribological parameters and the superior simulation speeds of hybrid NN-FVM frameworks compared to classical frameworks, several areas could be examined further in the future. For example, the effect of hybrid frameworks on simulations with different mesh sizes could be investigated to assess their scalability and robustness across different resolutions. Moreover, the performance of alternative neural networks or other machine learning tools in terms of their prediction accuracy could be investigated. The impact of the number of training samples on prediction accuracy could also be examined. Finally, there is the potential to improve prediction accuracy by training different ANNs for smaller regions of the parameter space or with a smaller number of input parameters. This way, each ANN can be trained with data that are more similar to each other, thus reducing the risk of bias towards a specific region of the parameter space or towards a specific type of output (e.g. profiles with large variations or large absolute magnitudes).

4. Summary and conclusions

Six different ANNs were trained to predict scalar values of rigid separation, maximum fluid temperature and mid-film friction coefficient, two-dimensional profiles of pressure and liquid film fraction and three-dimensional fluid temperature profiles in TEHL point contacts.

The first aim of this study was to test the performance of ML models in predicting the aforementioned parameters. The results showed that ANNs can predict scalar values, two-dimensional profiles as well as three-dimensional profiles very accurately. The rigid separation, maximum fluid temperature and mid-film friction coefficient predictions had R^2_{rs} values of 0.997, 0.985 and 0.939 respectively with a value of 1 indicating a perfect correlation between the predicted and true values. The pressure and liquid film fraction profile predictions were also very accurate both with mean R^2_{rs} values of 0.997 across the whole testing dataset. The mean R^2_{rs} for the fluid temperature profile prediction was smaller, with a value of 0.873 across the whole testing dataset. The reason for the smaller accuracy of the temperature profile predictions could be the fact that there is no empirical quantity to normalise the temperature profiles so that they are all similar in magnitude. Instead the temperature profiles were normalised by the *known* maximum temperature of each sample during training and then denormalised using

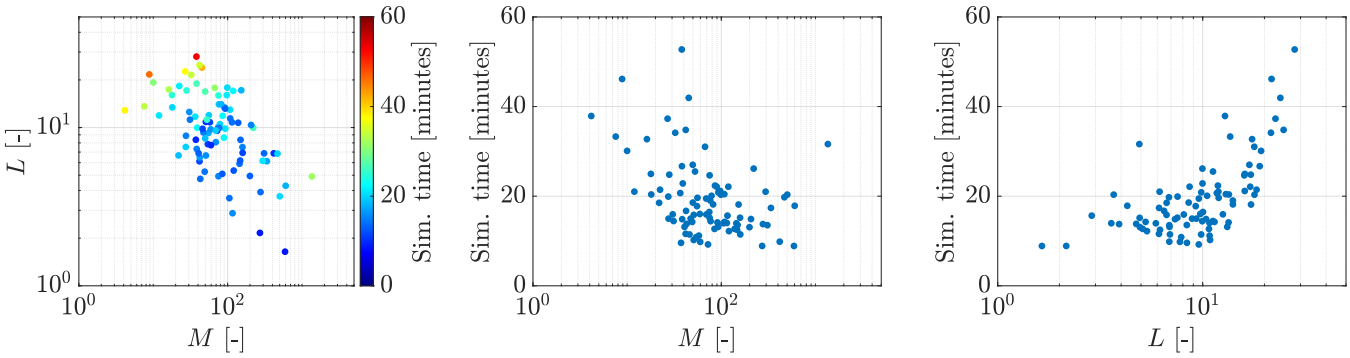


Figure 25: Simulation times of Classical FW simulations across parameter space (t_{mean} = 19.43 minutes).

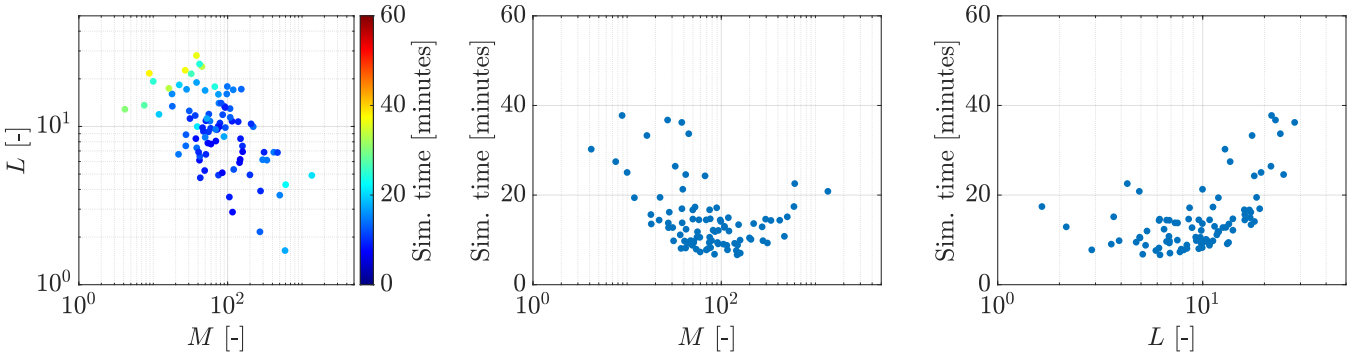


Figure 26: Simulation times of Hybrid FW 1 simulations across parameter space (t_{mean} = 14.42 minutes).

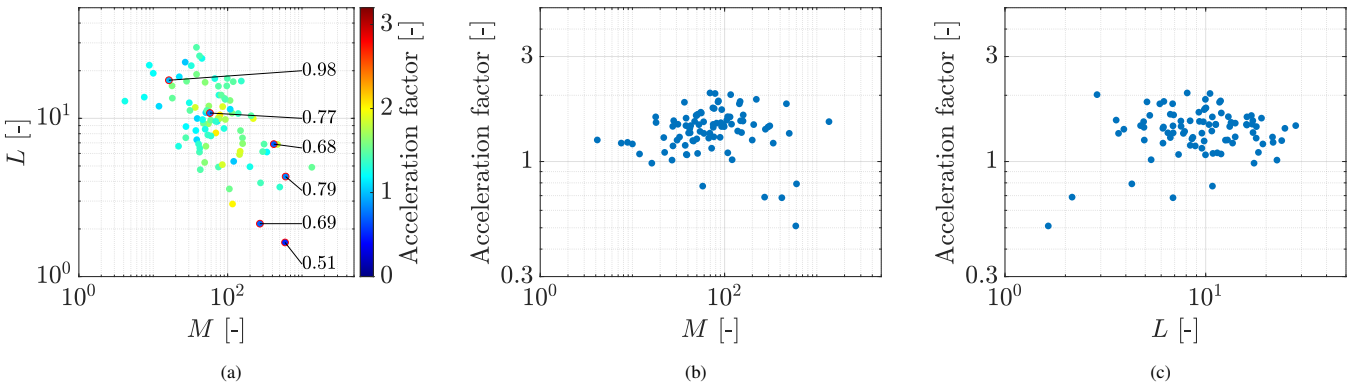


Figure 27: Acceleration factors of Hybrid FW 1 simulations across parameter space. The points outlined in red indicate an acceleration factor smaller than 1. The respective values are presented for clarity.

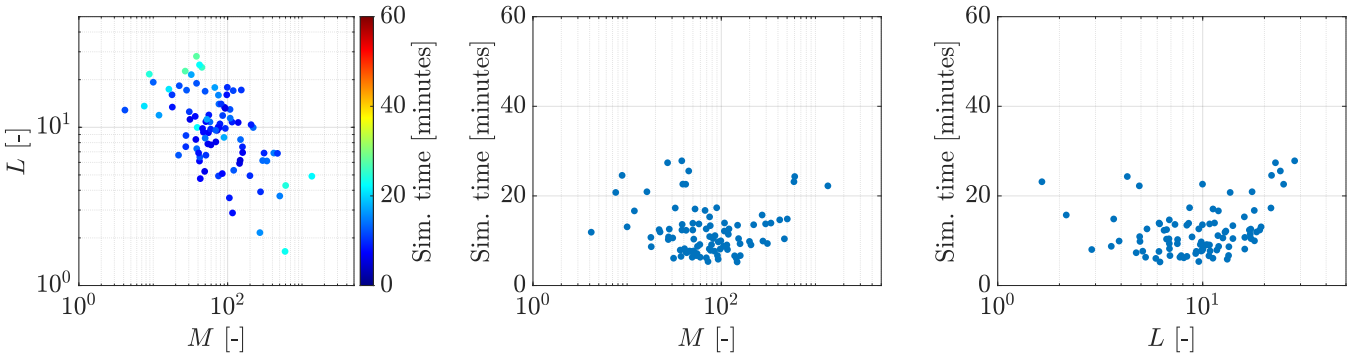


Figure 28: Simulation times of Hybrid FW 2 simulations across parameter space (t_{mean} = 12.09 minutes).

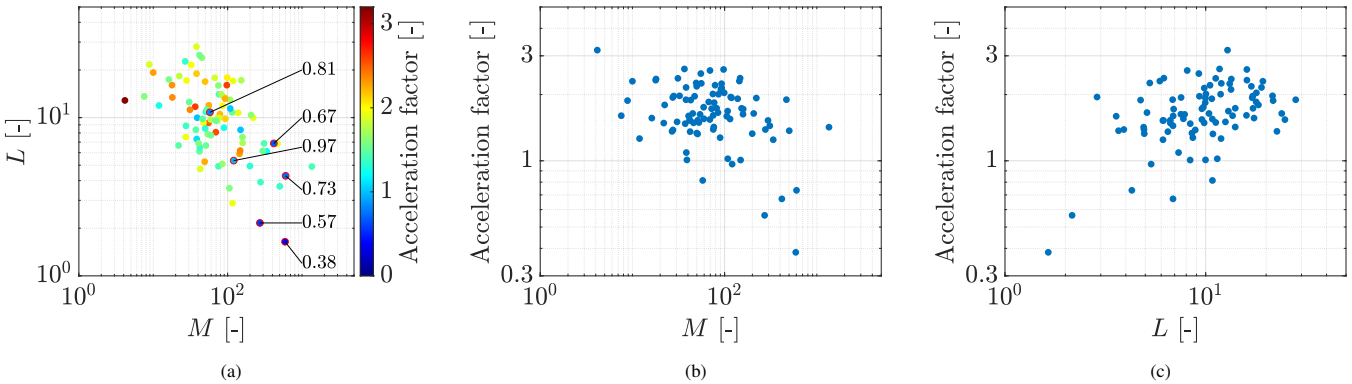


Figure 29: Acceleration factors of Hybrid FW 2 simulations across parameter space. The points outlined in red indicate an acceleration factor smaller than 1. The respective values are presented for clarity.

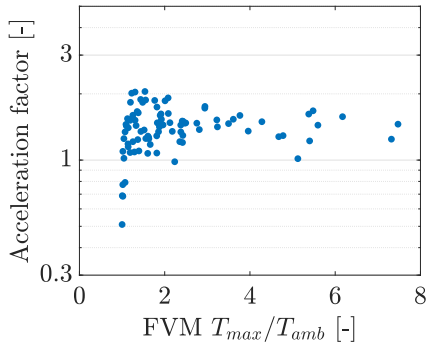


Figure 30: Effect of Classical FW normalised maximum temperature on acceleration factor (Hybrid FW 1).

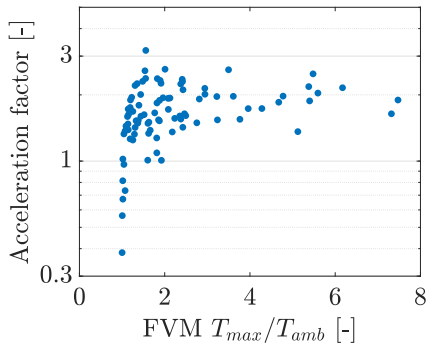


Figure 31: Effect of normalised maximum temperature on acceleration factor (Hybrid FW 2).

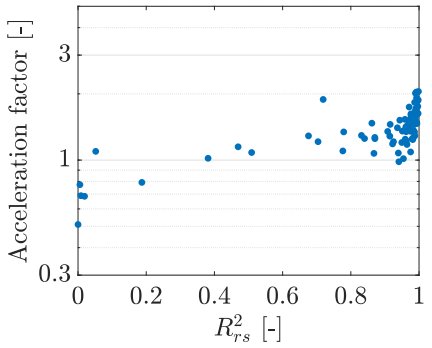


Figure 32: Effect of predicted temperature profile accuracy on acceleration factor (Hybrid FW 1).

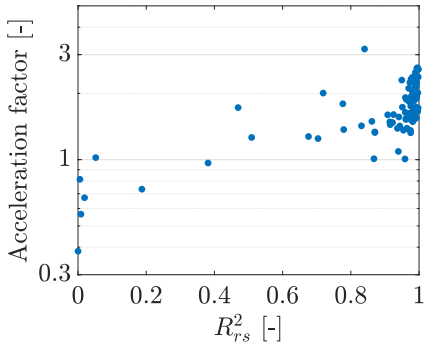


Figure 33: Effect of predicted temperature profile accuracy on acceleration factor (Hybrid FW 2).

the *predicted* maximum temperature during testing. This introduced some error in the temperature profile predictions, with the profiles with small temperature variations being affected the most from this error, thus being predicted the least accurately.

The second aim of this study was to test the speed of hybrid frameworks that combine ML models and classical finite volume approaches to solve TEHL problems. The results showed that the initial approximations used in classical TEHL simulations deviate significantly from the final solutions at high M and low L values in the case of rigid separation, at low M and high L values in the case of pressure and especially as L increases in the case of temperature, suggesting that these initial approximations can be greatly improved upon in terms of their match to the final solutions. Three frameworks were compared. The first framework (Classical FW) was a classical FVM framework, the second framework (Hybrid FW 1) was a hybrid NN-FVM framework with initial temperature prediction and the final framework (Hybrid FW 2) was a hybrid NN-FVM framework with initial rigid separation, pressure, liquid film fraction and fluid temperature prediction.

The results showed that the hybrid NN-FVM frameworks led to significant reduction in simulation times with the mean simulation time across the whole testing dataset using the Hybrid FW 1 being 14.42 minutes, 25.78% less than the mean simulation time using the Classical FW. The mean simulation time across the whole testing dataset using the Hybrid FW 2 was even shorter at 12.09 minutes, which was 37.78% less than the the mean simulation time using the Classical FW. Using the Hybrid FW 1, some simulations were up to twice as fast the Classical FW simulations, while using the Hybrid FW 2, some simulations were even three times as fast as the Classical FW simulations. Out of the 90 samples that were tested, the simulation times of six of those samples were actually longer when using the hybrid NN-FVM frameworks compared to the Classical FW. All six of those simulations were simulations with very small temperature variations and badly predicted temperature profiles. Simulation acceleration was linked to the accuracy of the temperature profile ANN predictions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Raw data used to plot the figures can be found online at: https://doi.org/10.5281/zenodo.15430178 (to be activated on acceptance).

Acknowledgements

F.K. thanks the UK Department of Science, Innovation and Technology (DSIT), the Engineering and Physical Sciences Research Council (EPSRC), and Shell for PhD funding through an iCASE studentship (EP/X524773/1). The authors thank Shell and the EPSRC for funding via the InFUSE Prosperity Partnership (EP/V038044/1). D.D. acknowledges the support of the Royal Academy of Engineering (RAEng) for the Shell/RAEng Research Chair in Complex Engineering Interfaces. J.P.E. acknowledges the support of the RAEng through their Research Fellowships scheme.

References

[1] A. Vakis, V. Yastrebov, J. Scheibert, C. Minfray, L. Nicola, D. Dini, A. Almqvist, M. Paggi, S. Lee, G. Limbert, J. Molinari, G. Anciaux, R. Aghababaei, S. E. Restrepo, A. Papangelo, A. Cammarata, P. Nicolini, C. Putignano, G. Carbone, S. Stupkiewicz, J. Lengiewicz, G. Costagliola, F. Bosia, R. Guarino, N. Pugno, M. Müser, M. Ciavarella, Modeling and simulation in tribology across scales: An overview, Tribology International 125 (2018) 169–199. doi:10.1016/j.triboint.2018.02.005.

[2] S. Ardah, F. J. Profito, D. Dini, A comprehensive review and trends in lubrication modelling, Advances in Colloid and Interface Science 342 (2025) 103492. doi:10.1016/j.cis.2025.103492.
URL https://doi.org/10.1016/j.cis.2025.103492

[3] P. M. Lugt, G. E. Morales-Espejel, A review of elasto-hydrodynamic lubrication theory, Tribology Transactions 54 (2011) 470–496. doi:10.1080/10402004.2010.551804.

[4] H. S. Cheng, A Refined Solution to the Thermal-Elastohydrodynamic Lubrication of Rolling and Sliding Cylinders, ASLE Transactions 8 (1965) 397–410. doi:10.1080/05698196508972110.

[5] A. W. Crook, The Lubrication of Rollers III. A Theoretical Discussion of Friction and the Temperatures in the Oil Film, Philosophical Transactions of the Royal Society A 254 (1961) 237–258. doi:10.1098/rsta.1961.0016.

[6] D. Zhu, S. zhu Wen, A Full Numerical Solution for the Thermoelastohydrodynamic Problem in Elliptical Contacts, Journal of Tribology 106 (1984) 246–254. doi:10.1115/1.3260895.

[7] X. Liu, M. Jiang, P. Yang, M. Kaneta, Non-Newtonian Thermal Analyses of Point EHL Contacts Using the Eyring Model, Journal of Tribology 127 (2005) 70–81. doi:10.1115/1.1843161.

[8] W. Habchi, D. Eyheramendy, S. Bair, P. Vergne, G. Morales-Espejel, Thermal elastohydrodynamic lubrication of point contacts using a Newtonian/generalized Newtonian lubricant, Tribology Letters 30 (2008) 41–52. doi:10.1007/s11249-008-9310-9.

[9] X. Wang, Y. Liu, D. Zhu, Numerical Solution of Mixed Thermal Elastohydrodynamic Lubrication in Point Contacts With Three-Dimensional Surface Roughness, Journal of Tribology 139 (2017). doi:10.1115/1.4032963.

[10] S. Liu, Q. Wang, G. Liu, A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses, Wear 243 (2000) 101–111. doi:10.1016/S0043-1648(00)00427-0.

[11] D. Zhu, On some aspects of numerical solutions of thin-film and mixed elastohydrodynamic lubrication, Proceedings of the Institution of Mechanical Engineers, Part J 221 (2007) 561–579. doi:10.1243/13506501JET259.

[12] S. Ardah, F. J. Profito, D. Dini, An integrated finite volume framework for thermal elasto-hydrodynamic lubrication, Tribology International 177 (2023) 107935. doi:10.1016/j.triboint.2022.107935.

[13] T. Lohner, A. Ziegltrum, J.-P. Stemplinger, K. Stahl, Engineering Software Solution for Thermal Elastohydrodynamic Lubrication Using Multiphysics Software, Advances in Tribology 2016 (2016) 1–13. doi:10.1155/2016/6507203.

[14] I. Pan, L. R. Mason, O. K. Matar, Data-centric Engineering: integrating simulation, machine learning and statistics. Challenges and opportunities, Chemical Engineering Science 249 (2022). doi:10.1016/j.ces.2021.117271.

[15] I. Argatov, Artificial Neural Networks (ANNs) as a Novel Modeling Technique in Tribology, Frontiers in Mechanical Engineering 5 (2019) 30. doi:10.3389/fmech.2019.00030.

[16] A. Rosenkranz, M. Marian, F. J. Profito, N. Aragon, R. Shah, The use of artificial intelligence in tribology-A perspective, Lubricants 9 (2021) 2. doi:10.3390/lubricants9010002.

[17] M. Marian, S. Tremmel, Current Trends and Applications of Machine Learning in Tribology-A Review, Lubricants 9 (2021) 86. doi:10.3390/lubricants9090086.

[18] A. T. Sose, S. Y. Joshi, L. K. Kunche, F. Wang, S. A. Deshmukh, A review of recent advances and applications of machine learning in tribology, Physical Chemistry Chemical Physics 25 (2023) 4408–4443. doi:10.1039/d2cp03692d.

[19] R. Shah, R. Jaramillo, G. Thomas, T. Rayhan, N. Hossain, M. Kchaou, F. J. Profito, A. Rosenkranz, Artificial Intelligence and Machine Learning in Tribology: Selected Case Studies and Overall Potential, Advanced Engineering Materials X (2025) 2401944. doi:10.1002/adem.202401944.

[20] M. Marian, J. Mursak, M. Bartz, F. J. Profito, A. Rosenkranz, S. Wartzack, Predicting EHL film thickness parameters by machine learning approaches, Friction 11 (2023) 992–1013. doi:10.1007/s40544-022-0641-6.

[21] W. Habchi, S. Bair, Machine-Learning-Assisted Identification and Formulation of High-Pressure Lubricant-Piezoviscous-Response Parameters for Minimum Film Thickness Determination in Elastohydrodynamic Circular Contacts, Tribology Letters 72 (2024) 134. doi:10.1007/s11249-024-01937-2.

[22] M. Tošić, M. Marian, W. Habchi, T. Lohner, K. Stahl, Application of machine learning for film thickness prediction in elliptical EHL contact with varying entrainment angle, Tribology International 199 (2024) 109940. doi:10.1016/j.triboint.2024.109940.

[23] J. Walker, H. Questa, A. Raman, M. Ahmed, M. Mohammadpour, S. R. Bewsher, G. Offner, Application of Tribological Artificial Neural Networks in Machine Elements, Tribology Letters 71 (2023) 3. doi:10.1007/s11249-022-01673-5.

[24] J. Issa, A. E. Hajj, P. Vergne, W. Habchi, Machine Learning for Film Thickness Prediction in Elastohydrodynamic Lubricated Elliptical Contacts, Lubricants 11 (2023) 497. doi:10.3390/lubricants11120497.
URL https://www.mdpi.com/2075-4442/11/12/497

[25] J. Kelley, V. Schneider, G. Poll, M. Marian, Enhancing practical modeling: A neural network approach for locally-resolved prediction of elastohydrodynamic line contacts, Tribology International 199 (2024). doi:10.1016/j.triboint.2024.109988.

[26] A. Singh, M. Wolf, G. Jacobs, F. König, Machine learning based surrogate modelling for the prediction of maximum contact temperature in EHL line contacts, Tribology International 179 (2023). doi:10.1016/j.triboint.2022.108166.

[27] S. Ardah, F. J. Profito, T. Reddyhoff, D. Dini, Advanced modelling of lubricated interfaces in general curvilinear grids, Tribology International 188 (2023) 108727. doi:10.1016/j.triboint.2023.108727.

[28] B. J. Hamrock, D. Dowson, Isothermal Elastohydrodynamic Lubrication of Point Contacts: Part III—Fully Flooded Results, Journal of Lubrication Technology 99 (1977) 264–275. doi:10.1115/1.3453074.

[29] C. J. A. Roelands, W. O. Winer, W. A. Wright, Correlational Aspects of the Viscosity-Temperature-Pressure Relationship of Lubricating Oils (Dr In dissertation at Technical University of Delft, 1966), Journal of Lubrication Technology 93 (1971) 209–210. doi:10.1115/1.3451519.

[30] H. Eyring, Viscosity, Plasticity, and Diffusion as Examples of Absolute Reaction Rates, The Journal of Chemical Physics 4 (1936) 283–291. doi:10.1063/1.1749836.

[31] D. Dowson, G. Higginson, Elasto-Hydrodynamic Lubrication, 1st Edition, Vol. 23, Pergamon Press, 1977. doi:10.1016/C2013-0-05764-7.

[32] P. Instruments, MTM (2025).
URL https://pcs-instruments.com/product/mtm/

[33] H. Spikes, Basics of EHL for practical application, Lubrication Science 27 (2015) 45–67. doi:10.1002/ls.1271.

[34] H. Moes, On Survey Diagrams, Tech. rep., Laboratory of Tribology, University of Twente, the Netherlands (1985).

[35] H. Moes, On Asymptotical Film Thickness Calculations, Tech. rep., Laboratory of Tribology, University of Twente, the Netherlands (1986).

[36] E. Guresen, G. Kayakutlu, Definition of artificial neural networks with comparison to other networks, Procedia Computer Science 3 (2011) 426–433. doi:10.1016/j.procs.2010.12.071.

[37] T. Szandała, Bio-inspired Neurocomputing. Studies in Computational Intelligence, Springer, 2021, Ch. Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks.

[38] D. P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, in: 3rd International Conference for Learning Representations, 2015. doi:10.48550/arXiv.1412.6980.

[39] T. Agrawal, Hyperparameter Optimization in Machine Learning, Apress, 2021. doi:10.1007/978-1-4842-6579-6.

[40] H. Moes, Optimum similarity analysis with applications to elastohydrodynamic lubrication, Wear 159 (1992) 57–66. doi:10.1016/0043-1648(92)90286-H.

[41] G. Nijenbanning, C. Venner, H. Moes, Film thickness in elastohydrodynamically lubricated elliptic contacts, Wear 176 (1994) 217–229. doi:10.1016/0043-1648(94)90150-3.