# Machine Learning-Driven Framework for Identifying Parameter-Driven Anomalies in Multiphysics Simulations

Zohreh Moradinia, Hans Vandierendonck and Adrian Murphy
*Queen's University Belfast, Belfast, U.K.*

Keywords: Multiphysics Simulations, Anomaly Detection, Machine Learning.

Abstract: This paper addresses the critical challenges associated with error management in multiphysics simulations, particularly regarding the sensitivity of these systems to parameter selection, which can lead to convergence failures and anomalies in simulation outputs. We propose a comprehensive analytical framework that systematically identifies the relationships between simulation parameters and governing equations, enabling the analysis of resulting anomalies. The framework classifies these anomalies, providing insights that inform the selection of appropriate unsupervised machine-learning algorithms for effective anomaly detection. To demonstrate the applicability of this approach, we apply the framework to a heat conjugate transfer (HCT) problem, integrating the heat transfer and Navier-Stokes equations. By thoroughly investigating parameter-driven anomalies, our framework enhances the reliability, convergence, and fidelity of multiphysics simulations, ultimately contributing to the robustness and accuracy of simulation outcomes.

## 1 INTRODUCTION

Multiphysics simulations are integral to a broad range of scientific and engineering applications, providing detailed insights into complex systems that involve interactions between multiple physical phenomena. Traditional modelling approaches, including numerical methods, analytical techniques, and equivalent circuit models, are widely employed in this field. Numerical methods such as the Finite Element Method (FEM) and Finite Volume Method (FVM) are renowned for their accuracy and robustness. However, their reliance on significant computational power, memory, and time resources poses considerable challenges, especially when simulating intricate or large-scale systems (Rinaldi, 2001; Peter H. Aaen and Balanis, 2006). In contrast, analytical methods are computationally efficient and precise but are typically restricted to simple geometries, limiting their applicability to more complex structures (Zhang, 2021). Similarly, equivalent circuit methods simplify computational complexity but are inefficient for novel devices requiring iterative adjustment (Junquan Chen and Xu, 2012).

The use of multiphysics simulations inherently involves managing various sources of error. Errors in these simulations can stem from modelling assumptions, numerical discretization, and finite-precision arithmetic, and they can significantly affect the accuracy and reliability of the results (Oberkampf and Trucano, 2002). Addressing these errors is critical, particularly in high-consequence applications where even small inaccuracies can have substantial impacts. Modelling errors arise from incomplete representations of the physical system, round-off errors result from limited numerical precision in computational arithmetic, and discretization and truncation errors are introduced during the discretization process of continuous equations (Heng Xiao and Roy, 2016; Tyson, 2018). These error sources necessitate careful mitigation strategies, such as grid refinement, model calibration, and higher-order discretization schemes, to improve simulation accuracy.

Despite the critical importance of minimizing errors, maximizing precision in all aspects of a simulation is often resource-intensive. Researchers frequently adopt a conservative approach, prioritizing maximum precision in parameter selection to ensure accuracy and result convergence. However, this practice often leads to prolonged computational times, even when high precision is not necessary for every stage of the simulation. The challenge of balancing accuracy with computational efficiency has been a persistent issue in the field, as researchers must often make trade-offs between simulation precision and performance (Committee, 1998; Christo-



Moradinia, Z., Vandierendonck, H. and Murphy, A.
Machine Learning-Driven Framework for Identifying Parameter-Driven Anomalies in Multiphysics Simulations.
DOI: 10.5220/0013514200003970
In *Proceedings of the 15th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2025)*, pages 278-286
ISBN: 978-989-758-759-7; ISSN: 2184-2841

pher J. Roy and Oberkampf, 2003; Christopher J. Roy and McWherter-Payne, 2003; He and Ding, 2001). This balance becomes even more complex when considering the uncertainty in input parameters, boundary conditions, and material properties, which can exacerbate the inherent uncertainty in multiphysics simulations.

Previous studies have proposed several approaches to address these challenges, including uncertainty quantification (UQ) methods and precision control techniques. UQ is essential for assessing how variability in model inputs influences simulation outputs, but traditional UQ techniques such as reduced-order modelling, polynomial chaos expansions, and Monte Carlo sampling are computationally expensive and often necessitate simplifying the model (R. A. Adams and Schmid, 2012; Ghanem and Spanos, 1991; Oakley, 2004). While these methods reduce the dimensionality or complexity of the problem, they may inadvertently limit the scope of the simulation or compromise its accuracy. Moreover, precision management strategies, such as adjusting arithmetic precision or discretization step sizes, have been explored to mitigate computational cost, but their effectiveness is constrained by the conflicting demands of accuracy and performance (Harvey and Verseghy, 2016; V. Chandola and Kumar, 2009).

To address these limitations, this paper introduces a novel approach leveraging machine learning (ML)-based anomaly detection techniques for multiphysics simulations. the proposed technique identifies anomalies—instances where simulation results deviate from expected outcomes—without altering the complexity of the simulation. This approach allows for the monitoring of simulation performance and the detection of critical points where errors accumulate or accuracy is compromised, effectively serving as an early warning system for simulation failures. The key contribution of this paper is a machine learning-based anomaly detection framework that enhances the accuracy and reliability of multiphysics simulations while reducing computational costs. This approach enables practitioners to make informed decisions regarding precision levels and parameter selection, thereby optimizing simulation performance without sacrificing accuracy. The proposed method enables the exploration of a broader range of simulation configurations while optimizing both precision and computational efficiency.

In this study, we apply the proposed anomaly detection technique to a heat conjugate transfer (HCT) problem, using heat transfer and Navier-Stokes equations to illustrate its effectiveness in identifying simulation anomalies. This method provides a more efficient and computationally feasible alternative to traditional error management strategies in multiphysics simulations. Moreover, the approach facilitates a deeper understanding of the trade-off between simulation precision and performance, enabling the selective adjustment of parameters based on specific simulation needs to avoid the common practice of uniformly applying maximum precision and unnecessarily resource-intensive.

## 2 METHOD

In this study, we present a framework for detecting anomalies in multiphysics simulation results by analyzing the relationship between effective parameters and the governing physical equations. Our approach integrates multiphysics simulations with ML anomaly detection algorithms. Specifically, we apply this method to the conjugate heat transfer problem, focusing on flow over a heated plate. Open-source solvers and coupling tools are utilized to conduct the simulations. The methodology consists of three key steps. First, we assess the influence of various parameters—including physical, material, and simulation parameters—on the simulation outcomes. Understanding how these parameters affect the results is crucial for identifying potential anomalies. Second, we investigate the relationship between these parameters and the governing equations to determine whether improper parameter settings could adversely impact the equations and, consequently, the simulation results. This analysis enables us to quantify the extent to which incorrect parameter values contribute to deviations from expected outcomes. Finally, based on the insights gained from the parameter-equation relationship, we select an appropriate ML anomaly detection algorithm. A key requirement for the algorithm is that it should not rely on predefined labeled data, as anomalies in simulation results can manifest in diverse ways. Therefore, we employ unsupervised learning algorithms, which do not require prior training on labeled datasets and making it ideal for rare and unknown anomalies. Additionally, unsupervised methods can identify previously unseen anomalies by learning normal system behavior and effectively scale to the large datasets typical of multiphysics problems. Many unsupervised algorithms are also computationally efficient, making them suitable for real-time or iterative simulations.However, selecting the most suitable unsupervised algorithm depends on the specific multiphysics problem, as it is influenced by both the governing equations and the associated parameters. To verify the model, we compare its outcomes when applied to simulation results obtained under different configurations. Specifically, we examine cases where the value of one of the effective parameters differs from the default configuration settings and compare these results with the ground truth obtained from the default simulation setup. This comparison allows us to assess the model's ability to detect anomalies and validate its accuracy in identifying deviations caused by parameter variations. Ultimately, this framework enables the development of a model that assists in configuring simulations appropriately for a given application. It provides valuable insights into simulation results, facilitating the identification of anomalies specific to the problem. Moreover, it aids in optimizing the balance between computational speed and accuracy based on application requirements, enhancing the overall efficiency and reliability of multiphysics simulations.

# 3 CASE STUDY

Conjugate heat transfer (CHT)(M. Vynnycky and Pop, 1998) is a critical phenomenon that involves the transfer of heat between fluids and solid boundaries, where heat passes from one fluid to the solid boundary and then transfers from the solid to the other fluid. This process is particularly important in engineering applications such as heat exchanger design and cooling systems for electronic components.

In this study, we focus on a two-dimensional CHT problem involving a rectangular, thermally conducting slab with laminar, incompressible fluid flow over it. This scenario presents a classic CHT challenge, where the steady heat transfer dynamics must be accurately captured to ensure reliable simulation results. The forced flow occupies the region $-\infty \leq x \leq \infty, y > 0$, with uniform velocity $U_s$ and temperature $T$, while the conducting slab occupies the region $-a \leq y \leq 0$, $-\frac{b}{2} \leq x \leq \frac{b}{2}$. A schematic of this setup is illustrated in Figure 1. To accurately model this problem, two governing equations are used: the Navier-Stokes equation for fluid motion and the heat transfer equation for

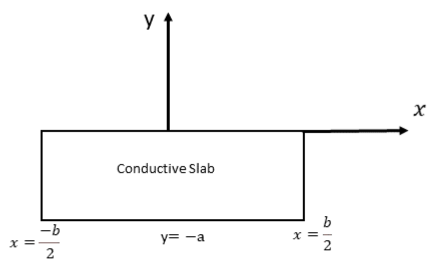


Figure 1: Physical model and coordinate system of flow over heated plate.

temperature distribution. The Navier-Stokes equation governs the motion of the fluid and can be written as:

$$\rho \frac{Dv}{Dt} = -\nabla p + \mu \nabla^2 v + f, \qquad (1)$$

Where $\rho$ is the fluid density, $v$ is the velocity, $p$ is the pressure, and $\mu$ is the fluid viscosity. This equation accounts for both convection and diffusion in the fluid's motion. The heat transfer equation, which governs the temperature distribution in the system, is expressed as:

$$\rho c_p \frac{DT}{Dt} = \nabla \cdot (k \nabla T) + Q, \qquad (2)$$

where $\rho$ is the density, $c_p$ is the specific heat capacity at constant pressure, $\frac{DT}{Dt}$ is the substantial (material) derivative of temperature, $k$ is the thermal conductivity, and $Q$ is the volumetric heat generation rate.

To simulate fluid flow and heat transfer, we utilize OpenFOAM (H. Jasak, 2007)software, employing the laplacianFoam solver for steady-state heat conduction in solid slabs and the buoyantPimpleFoam solver for buoyancy-driven fluid flow and transient simulations using the PIMPLE algorithm. Accurate coupling of solvers for fluid and solid domains in CHT simulations is addressed with PreCICE software(H. J. Ungartz, 2016), which ensures consistent heat flux and temperature transfer between regions, enhances accuracy at the solid-fluid interface, and improves performance through iterative methods, effective solver communication, and data mapping for non-matching grids. This multiphysics problem analyzes heat flux and temperature exchange to determine how key parameters—including time step, convergence limitation, CP, Reynolds number, and Prandtl number—impact the heat and Navier-Stokes equations, identifying simulation discontinuities, as outlined in Table 1.

# 4 STUDING EFFECTIVE PARAMETERS

In managing the precision and computational efficiency of simulations, there is a tendency to avoid configuring all parameters at the highest precision due to the significant increase in computational cost and execution time. However, when parameters are not set to these high-precision values, anomalies may emerge in the simulation results. Figures 2 illustrate the complex relationship between time steps and heat capacity in relation to the output parameters of the simulation. Each data point on these graphs corresponds to a unique simulation configuration. In this study, errors are defined as the mean difference between the simulation outputs for a given parameter configuration and

Table 1: Definition of studied Parameters in HCT problem.

| **Parameter** | **Definition** |
|---|---|
| Time Step | The discrete intervals for simulation progression, affecting accuracy and stability. |
| Convergence Limitation | A threshold that indicates when the iterative solution has achieved acceptable accuracy based on differences between iterations. |
| CP | The heat required to change a unit mass of a material by one degree Celsius, affecting its temperature response and heat transfer. |
| Reynolds Number | A dimensionless parameter that classifies fluid flow as laminar or turbulent based on inertial versus viscous forces. |
| Prandtl Number | A dimensionless ratio that indicates the relative importance of momentum transfer to heat transfer in fluid flows. |

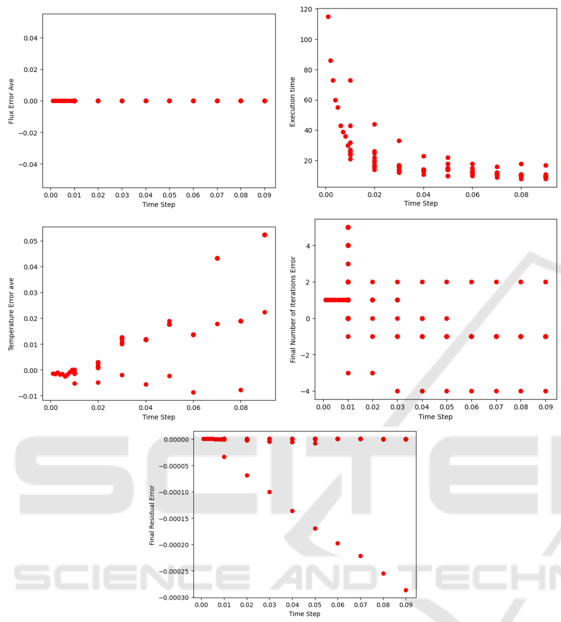

Figure 2: Time-steps versus selected outputs in flow over heated plate problem- average error of temperature and flux, final residual, iteration numbers, and execution time

those obtained using the default configuration. Figure 2 highlights the minimal error range observed for both the average temperature and the final residual values, with negligible error in the flux average across all time steps. Notably, the uniformity in the number of iterations across different time steps underscores the robustness of the results. Importantly, our findings indicate that increasing the time step size results in a nearly 50% reduction in execution time while maintaining high simulation accuracy, with errors below 0.05% for the temperature output.

These results suggest that a wide range of parameter configurations can be used without compromising simulation precision, enabling faster simulations. The insights gained from our findings help to elucidate the complexities involved in multiphysics simulations, revealing important relationships between critical parameters and their outcomes. The flexibility to adjust parameter values, while it may slightly impact accuracy, is crucial in optimizing simulations for efficiency.

# 5 ANALYSIS OF PARAMETER IN HEAT TRANSFER AND NAVIER-STOKES EQUATIONS

This section explores the influence of the Reynolds number (Re), Prandtl number (Pr), CP, and numerical modelling parameters such as time step and convergence limitations on the solutions of the heat transfer and Navier-stokes equations. We aim to understand their potential to create or smooth out discontinuities in the temperature field and heat flux.

### 5.0.1 Non-Dimensional Form of Heat Transfer Equations

To understand the effects of the $Re$, $Pr$, and $c_p$, we non-dimensionalize the heat transfer equation. below are dimensionless variables:

$$T^* = \frac{T - T_\infty}{\Delta T}, \quad \mathbf{u}^* = \frac{\mathbf{u}}{U}, \quad x^* = \frac{x}{L}, \quad t^* = \frac{tU}{L}, \tag{3}$$

where: $U$ is a characteristic velocity, $L$ is a characteristic length, $T_\infty$ is the reference temperature, $\Delta T$ is the temperature difference. Substituting these into the heat conduction equation, we first need to express all terms in terms of the dimensionless variables. Start with the substantial derivative term. Substitute the dimensionless variables:

$$T = T_\infty + T^*\Delta T, \mathbf{u} = U\mathbf{u}^*, x = Lx^*, t = \frac{L}{U}t^*. \tag{4}$$

This is the non-dimensional form of the heat conduction equation:

$$\frac{\partial T}{\partial t} + \mathbf{u}\cdot\nabla T = \frac{1}{Re\cdot Pr}\nabla^2 T + \frac{QL}{\rho c_p U \Delta T}. \tag{5}$$

### 5.0.2 Non-Dimensionalization of the Navier-Stokes Equations

The non-dimensionalization process can be applied to the Navier-Stokes equations to better understand how various parameters influence heat flux.

$$\mathbf{u}^* = \frac{\mathbf{u}}{U}, \quad P^* = \frac{P}{\rho U^2}, \quad \mathbf{x}^* = \frac{\mathbf{x}}{L}, \quad t^* = \frac{tU}{L} \tag{6}$$

Where $U$ is a characteristic velocity, $L$ is a characteristic length, $P$ represents pressure, $\rho$ is density. After substituting and rearranging, the Navier-Stokes equations in their dimensionless form can be expressed as:

$$\frac{\partial \mathbf{u}^*}{\partial t^*} + \mathbf{u}^* \cdot \nabla^* \mathbf{u}^* = -\nabla^* P^* + \frac{1}{Re} \nabla^{*2} \mathbf{u}^* + \mathbf{F}^* \tag{7}$$

Here, $\mathbf{F}^*$ represents body forces expressed in non-dimensional form.

## 5.1 Effects of Physical Parameters in Heat Transfer and Navier-Stokes Equations

The analysis of temperature and heat flux can be significantly affected by physical parameters. Understanding these parameters is crucial for detecting anomalies caused by not incorrect values of these parameters

### 5.1.1 Reynolds Number (*Re*)

*Re* is a dimensionless quantity that characterizes the flow regime of a fluid, representing the ratio of inertial forces to viscous forces. It is defined as:

$$Re = \frac{\rho U L}{\mu} \tag{8}$$

where $\rho$ is the fluid density, $U$ is the characteristic velocity, $L$ is the characteristic length, and $\mu$ is the dynamic viscosity. In high Reynolds number scenarios, the convective term $\mathbf{u} \cdot \nabla T$ dominates over the diffusive term $\frac{1}{Re \cdot Pr} \nabla^2 T$, leading to stronger convective heat transfer and sharper temperature gradients, especially near walls (thermal boundary layers). Turbulence enhances mixing, smoothing temperature discontinuities in the bulk flow while maintaining rapid temperature changes near boundaries.

At low Reynolds numbers, viscous and thermal diffusion become more significant, with the term $\frac{1}{Re \cdot Pr} \nabla^2 T$ being larger. This enhances thermal diffusion, smoothing temperature gradients and resulting in a more orderly flow with smoother temperature distributions.

In regions with high shear, such as near solid boundaries, substantial velocity gradients can arise, leading to abrupt changes in the velocity field, creating potential discontinuities and turbulence.

### 5.1.2 Prandtl Number (*Pr*)

Pr is another essential dimensionless quantity that relates the momentum diffusivity (kinematic viscosity) to thermal diffusivity, defined as:

$$Pr = \frac{\nu}{\alpha} \tag{9}$$

where $\nu$ is the kinematic viscosity and $\alpha$ is the thermal diffusivity. In high $Pr$ scenarios, momentum diffusivity is significantly greater than thermal diffusivity. The term $\frac{1}{Re \cdot Pr}$ becomes small, making the thermal diffusion term $\nabla^2 T$ less influential. Consequently, the temperature gradient near surfaces can be steep, leading to pronounced thermal gradients and potential discontinuities. The Navier-Stokes equations can be approximated as:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{Re} \nabla P + \frac{1}{Re} \nabla^2 \mathbf{u} \tag{10}$$

The reduced thermal diffusion in high $Pr$ conditions can result in prolonged high temperature gradients, which can induce potential discontinuities near solid boundaries or within flows exhibiting significant thermal stratification. Conversely, when $Pr$ is low, thermal diffusivity becomes more effective, leading to a thicker thermal boundary layer and smoother temperature distributions. In this regime, the thermal diffusion term $\nabla^2 T$ is more significant, enhancing the conduction effect and reducing the likelihood of temperature discontinuities.

### 5.1.3 Specific Heat Capacity ($c_p$)

The specific heat capacity ($c_p$) plays a pivotal role in determining the thermal response of a fluid, defining the amount of heat required to change the temperature of a unit mass by one degree Celsius. A higher $c_p$ indicates that the fluid can absorb more heat before experiencing significant temperature changes, smoothing out temperature gradients. This increased thermal inertia also raises the thermal time constant:

$$\tau = \frac{\rho c_p L}{k} \tag{11}$$

Conversely, a low $c_p$ allows for a more rapid temperature increase in response to heat inputs, which can create sharper temperature gradients and increase the likelihood of discontinuities, particularly in regions with high heat flux. Although $c_p$ does not explicitly appear in the Navier-Stokes equations, it influences the coupled energy equation: This equation shows that a higher $c_p$ increases the thermal inertia of the fluid, resulting in smoother temperature variations and reducing discontinuities in both velocity and temperature fields, as the fluid takes longer to respond to thermal changes.

## 5.2 Numerical Modeling Considerations

### 5.2.1 Time Step ($\Delta t$)

The time step $\Delta t$ in numerical simulations affects both stability and accuracy. A smaller $\Delta t$ improves the resolution of temporal changes, capturing transient phenomena and reducing numerical artifacts, leading to more accurate and stable solutions, though at higher computational cost. The Courant-Friedrichs-Lewy (CFL) condition, $\Delta t < \frac{L}{U}$, ensures stability by limiting $\Delta t$ based on grid size and velocity, preventing instabilities and artificial discontinuities. Conversely, a larger $\Delta t$ can cause numerical instability, leading to oscillations, divergence, and errors in the temperature field. Violating the CFL condition with a large time step can result in unphysical results and the propagation of numerical errors.

### 5.2.2 Convergence Limitation

Numerical convergence is vital for accurate solutions, and issues can arise from inadequate meshing, improper boundary conditions, or insufficient iterations, leading to errors or discontinuities in the temperature field. A fine mesh resolves small-scale features, with the mesh size $\Delta x$ satisfying $\Delta x < \frac{L}{N}$ for $N$ grid points per characteristic length $L$, while insufficient meshing can cause aliasing and artificial discontinuities. Proper boundary conditions are essential to avoid unphysical solutions, and they must match the physical problem and be applied consistently to prevent spurious gradients. Sufficient iterations are also necessary for solvers to converge accurately, and meeting convergence criteria like residual norms ensures stable, accurate solutions. Incomplete convergence leaves unresolved gradients and potential discontinuities.

# 6 ANOMALY DETECTION FOR HCT

Based on the equations governing the parameters studied in Section 6, setting parameters beyond acceptable limits may cause the convective term $\mathbf{u} \cdot \nabla T$ to dominate over the diffusive term $\frac{1}{Re \cdot Pr} \nabla^2 T$. This dominance leads to stronger convective heat transfer and sharper temperature gradients. When the term $\frac{1}{Re \cdot Pr}$ becomes small, the thermal diffusion term $\nabla^2 T$ loses influence, resulting in steep temperature gradients near surfaces, which can cause significant thermal gradients and potential discontinuities.

When parameter values exceed the recommended range, the dominant terms in the governing equations can lead to anomalies and discontinuities in the simulation results.

To avoid surpassing acceptable error thresholds, limitations, or introducing discontinuities in the simulation results, we incorporate an additional step of anomaly detection.

Specifically, we evaluate the effectiveness of three anomaly detection algorithms—One-Class SVM, Isolation Forest, and LOF to identify anomalies in temperature and heat flux data generated from the simulations. These algorithms provide an important layer of analysis to detect deviations that could indicate emerging errors or inconsistencies in the results, helping to maintain the integrity of the simulations.

## 6.1 Local Outlier Factor

The LOF (M. M. Breunig and Sander, 2000) is an unsupervised anomaly detection algorithm that identifies anomalies by assessing local density deviations of data points relative to their neighbors. It is particularly effective in datasets with heterogeneous density distributions, where the density of data points varies significantly across different regions.

The algorithm first computes the k-distance for each data point, defined as the distance to its k-th nearest neighbor. The choice of $k$ critically influences the sensitivity to anomalies; smaller $k$ captures fine anomalies, while larger $k$ identifies broader patterns.

LOF computes reachability distances, which are defined for points $A$ and $B$ as the maximum of the actual distance between $A$ and $B$ and the k-distance of point $B$. This calculation ensures local density is accurately represented.

The LRD for each point is derived as the inverse of the average reachability distance to its k-nearest neighbors, reflecting the density of surrounding points.

LOF scores are calculated as the average ratio of the LRD of a point's neighbors to its own LRD. A score near 1 indicates a normal point, while a score significantly greater than 1 indicates an anomaly.

## 6.2 One-Class SVM

The One-Class SVM(K.-L. Li and Xu, 2003) is an unsupervised machine learning algorithm specifically designed for anomaly detection. It is utilized to identify data points that significantly deviate from the normal data distribution. Unlike conventional SVMs, which are primarily used for binary classification, the One-Class SVM is trained exclusively on normal data. Its goal is to construct a decision boundary that encloses the majority of the normal data points, allowing the detection of anomalies that lie outside this boundary.

SVM operates by receiving a dataset consisting solely of normal data points, along with a kernel function that transforms the input data into a higher-dimensional space. This transformation is key to enhancing the algorithm's ability to separate normal data points from potential anomalies, as it enables the construction of a hyperplane in this transformed space.

Once the decision function is learned based on the normal data, it is applied to classify new data points. Points that lie outside the established decision boundary are classified as anomalies, while those within the boundary are considered normal. Each new data point is assigned an anomaly score, where negative values denote normal points, and positive values indicate anomalies.

### 6.3 Isolation Forest

Isolation Forest(F. T. Liu and Zhou, 2008) is an unsupervised machine learning algorithm designed for anomaly detection by recursively partitioning the data. The underlying principle of Isolation Forest is that anomalies are "few and different," making them more susceptible to isolation than normal data points. By constructing random decision trees, the algorithm isolates individual data points and measures the path length from the root to the point as an indicator of normality: shorter paths suggest anomalies, while longer paths correspond to normal points.

Isolation Forest receives a dataset consisting of $N$ data points, along with predefined parameters such as the maximum tree height $h$ and the number of trees $T$ to be generated. It begins by constructing multiple isolation trees, each of which recursively partitions the data by selecting random features and random split values within the feature's range. The process continues until each data point is isolated or the maximum tree height is reached.

For each data point, the path length in each tree is computed, representing the number of splits required to isolate that point. By averaging the path lengths across all trees, the algorithm assigns an anomaly score to each point. Data points that are isolated quickly (i.e., have shorter path lengths) are more likely to be anomalies, while those requiring longer path lengths are more likely to be normal.

In this study, we employed several algorithms to detect discontinuities in heat transfer simulations caused by numerical errors, parameter variations, or physical phenomena. Both One-Class SVM and Isolation Forest effectively manage high-dimensional spaces, which is crucial for the complexity of heat transfer simulations involving multiple interacting parameters. Additionally, the LOF excels in identifying anomalies within datasets characterized by varying local densities, a common occurrence in heat transfer simulations where temperature and heat flux can differ significantly across regions. Furthermore, Isolation Forest is notably efficient and capable of handling large-scale simulations, making it well-suited for real-time anomaly detection in extensive heat transfer datasets. Overall, these unsupervised learning methods improve the detection of discontinuities, enhancing the reliability and accuracy of multiphysics simulations. Their capabilities in managing high-dimensional data, accommodating local density variations, and processing large datasets render them valuable tools for anomaly detection in complex engineering problems.

## 7 RESULTS AND DISCUSSION

In this section, we sought to identify the most effective method for detecting discontinuities in the simulation results. Each algorithm was assessed based on its ability to detect anomalies without generating excessive false positives, particularly in the context of subtle parameter variations. The simulation outputs using default parameters are presented in Figure 3. The temperature and flux data exhibit smooth and continuous behaviour in the plots, with no evidence of discontinuities or abrupt changes in the baseline truth data simulation. It is essential to identify any discontinuities or sharp transitions that may arise due to variations in the studied parameter values using algorithmic detection methods. Anomalies in simulation results associated with different configurations have been identified based on these analyses.

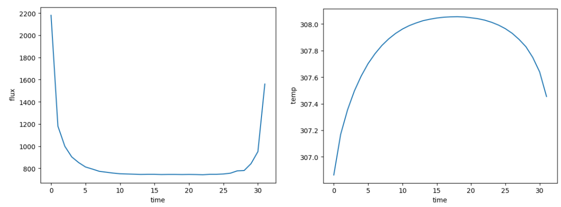


Figure 3: Temperature and Heat flux plots of flow over a heated plate simulation.

Figure 4 illustrates anomaly detection across three algorithms and each subplot represents variations in parameters with temperature and heat flux data. The black line represents the baseline, while anomalies are marked: blue squares for IF, green circles for SVM, and red crosses for LOF.

1. Anomaly Detection in Temperature Data: The right column of plots shows temperature over time for each varied parameter, generally following an initial rise before stabilizing, indicating thermal equilibration. This pattern helps evaluate each algorithm's ability to detect anomalies during both transient and steady-state phases. The IF algorithm detects a few anomalies, especially during the early rise phase, suggesting it is more attuned to significant outliers than to smaller deviations. SVM detects more anomalies overall, reflecting higher sensitivity to variations, though this may increase false positives, especially during a steady state. LOF, using local density, effectively identifies anomalies in the early, high-variance phase, making it valuable for rapid-change regions, though its detections decrease as the system stabilizes.

2. Anomaly Detection in Heat Flux Data: The left column shows heat flux over time for each parameter. Unlike temperature, heat flux spikes initially, then decays to a steady state, reflecting the system's thermal gradient settling into equilibrium.

IF detects a few anomalies, mainly during the initial spike, focusing on significant deviations and overlooking minor fluctuations in the stable phase. SVM detects anomalies consistently, including in stable periods, highlighting its sensitivity but with the potential for false positives as flux stabilizes. LOF captures a high density of anomalies in the initial, rapid-change phase but fewer as flux reaches a steady state, showing its strength in non-equilibrium conditions.

Parameter variations affect each algorithm uniquely. Changes in Cp strongly impact temperature and heat flux, yielding high anomaly counts across all algorithms. In contrast, Re primarily affects inertial forces, resulting in fewer detections. Ts and Pr, which directly influence thermal gradients, prompt higher sensitivity: Ts variation causes rapid temperature and flux spikes, detected well by LOF and SVM, while Pr changes create anomalies during the transition to a steady state, effectively captured by IF and LOF.

The comparative analysis across various parameter changes highlights the critical balance between sensitivity and specificity in anomaly detection algorithms. One-Class SVM and Isolation Forest, despite their capabilities, suffer from over-detection issues that compromise their reliability, particularly when variations are present. Conversely, LOF demonstrated consistent accuracy and robustness in identifying meaningful anomalies, suggesting it is the preferred method for ensuring the integrity of simulation results in the presence of parameter variations.

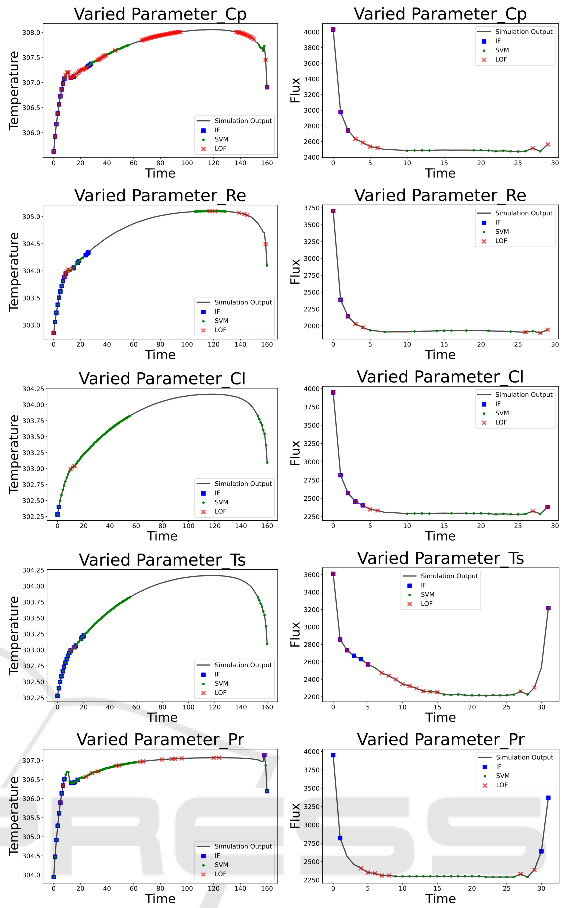


Figure 4: Anomaly detection results of Isolation Forest, one class SVM and LOF on Heat Flux and Temperature quantity of flow over a heated plate when selected parameters changed.

# 8 CONCLUSION

The challenges associated with managing accuracy and computational efficiency in multiphysics simulations have long been a focus of research. The necessity to balance precision and performance has underscored the demand for innovative solutions. This paper introduces a novel machine learning-based anomaly detection framework that enhances the reliability of simulations by identifying critical points where errors may arise. The key strength of our proposed approach lies in its ability to optimize simulation performance while maintaining accuracy, providing insights into the trade-off between precision and computational cost. Our results demonstrate the effectiveness of this approach in heat transfer problems, showcasing its versatility and potential applicability across various scenarios. Importantly, this methodology simplifies the error management process, significantly reducing computational volume while ensuring the integrity of the simulation outcomes. This framework not only enhances the efficiency of multiphysics simulations but also establishes a solid foundation for confidently determining optimal precision requirements, marking a meaningful advancement in the field.

## REFERENCES


Christopher J. Roy, M. A. M.-P. and Oberkampf, W. L. (2003). Verification and validation for laminar hypersonic flowfields, part 1: Verification. *AIAA Journal*, 41(10):1934–1943.

Christopher J. Roy, W. L. O. and McWherter-Payne, M. A. (2003). Verification and validation for laminar hypersonic flowfields, part 2: Validation. *AIAA Journal*, 41(10):1944–1954.

Committee, C. F. D. (1998). *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations (AIAA G-077-1998 (2002))*. American Institute of Aeronautics and Astronautics, Inc.

F. T. Liu, K. M. T. and Zhou, Z.-H. (2008). Isolation forest. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE.

Ghanem, R. and Spanos, P. (1991). Polynomial chaos in stochastic finite elements. *Journal of Applied Mechanics*, 57(1):197–205.

H. J. Ungartz, F. Lindner, B. G.-e. a. (2016). precice—a fully parallel library for multi-physics surface coupling. *Computers & Fluids*, 141:250–258.

H. Jasak, A. Jemcov, Z. T. e. a. (2007). Openfoam: A c++ library for complex physics simulations. In *International Workshop on Coupled Methods in Numerical Dynamics*, pages 1–20.

Harvey, R. and Verseghy, D. L. (2016). The reliability of single precision computations in the simulation of deep soil heat diffusion in a land surface model. *Climate Dynamics*, 46:3865–3882.

He, Y. and Ding, C. H. (2001). Using accurate arithmetics to improve numerical reproducibility and stability in parallel applications. *The Journal of Supercomputing*, 18:259–277.

Heng Xiao, J.-L. Wu, J.-X. W. R. S. and Roy, C. J. (2016). Quantifying and reducing model-form uncertainties in reynolds-averaged navier–stokes simulations: A data-driven, physics-informed bayesian approach. *Journal of Computational Physics*, 324:115–136.

Jun-quan Chen, Xing Chen, C.-J. L. K. H. and Xu, X.-B. (2012). Analysis of temperature effect on pin diode circuits by a multiphysics and circuit cosimulation algorithm. *IEEE Transactions on Electron Devices*, 59(11):3069–3077.

K.-L. Li, H.-K. Huang, S.-F. T. and Xu, W. (2003). Improving one-class svm for anomaly detection. In *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693)*, volume 5, pages 3077–3081. IEEE.

M. M. Breunig, H. P. Kriegel, R. T. N. and Sander, J. (2000). Lof: Identifying density-based local outliers. *ACM SIGMOD Record*, 29(2):93–104.

M. Vynnycky, S. Kimura, K. K. and Pop, I. (1998). Forced convection heat transfer from a flat plate: The conjugate problem. *International Journal of Heat and Mass Transfer*, 41(1):45–59.

Oakley, J. E. (2004). Using latin hypercube sampling in uncertainty analysis of complex models. In *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pages 3275–3280.

Oberkampf, W. L. and Trucano, T. G. (2002). Verification and validation in computational fluid dynamics. *Progress in Aerospace Sciences*, 38(3):209–272.

Peter H. Aaen, J. A. P. and Balanis, C. A. (2006). Modeling techniques suitable for cad-based design of internal matching networks of high-power rf/microwave transistors. *IEEE Transactions on Microwave Theory and Techniques*, 54(7):3052–3059.

R. A. Adams, D. J. I. and Schmid, C. T. (2012). Reduced order modelling for simulations of nonlinear dynamical systems. *SIAM Journal on Scientific Computing*, 34(3):A1273–A1296.

Rinaldi, N. (2001). On the modeling of the transient thermal behavior of semiconductor devices. *IEEE Transactions on Electron Devices*, 48(12):2796–2802.

Tyson, W. C. (2018). *On numerical error estimation for the finite-volume method with an application to computational fluid dynamics*. PhD thesis, Virginia Tech.

V. Chandola, A. B. and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):1–58.

Zhang, Y. (2021). Improved numerical-analytical thermal modeling method of the pcb with considering radiation heat transfer and calculation of components' temperature. *IEEE Access*, 9:92925–92940.