Digital twins of thermal systems: a comparison between supervised and reinforcement learning

Armando Di Meglio[1,*], **Nicola Massarotti**[1] **and Perumal Nithiarasu**[2]

[1] Department of Engineering, University of Naples "Parthenope", Centro Direzionale, Isola C4 – 80143, Italy
[2] Zienkiewicz Institute for Modelling, Data and AI, Swansea University, Swansea, SA1 8EN, United Kingdom

[*] e-mail: armando.dimeglio001@studenti.uniparthenope.it

Abstract. This article explores two novel approaches for controlling heat transfer systems through the development of "digital twins", focusing on transient thermal systems. The study involves creating a digital representation of a physical system, specifically a 2D square subjected to an inward and transient heat flux, with the goal of keeping the maximum temperature within a predefined limit by changing the convective cooling. The first method utilizes a neural network trained on steady-state data, whereas the second employs an interactive learning algorithm. Results show that both strategies prove to be effective in managing the system's thermal performance. However, the RL-based approach demonstrates greater flexibility in adapting to new scenarios, albeit at the cost of increased computational demands due to the necessity of integrating interactive learning with unsteady Finite Element Method (FEM) simulations for training, validation, and testing phases.

Keywords: *Digital Twin, heat transfer, Interactive AI, Deep Learning, Machine Learning, FEM simulation*

1. Introduction

In recent years, the intersection of Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), and the concept of Digital Twins (DT) has heralded a new era of innovation across various domains. AI refers to the simulation of human intelligence in machines programmed to mimic cognitive functions such as learning and problem-solving. ML, a subset of AI, empowers systems to automatically learn and improve from experience without being explicitly programmed. DL, in turn, leverages neural networks with many layers to learn representations of data with multiple levels of abstraction. A Digital Twin is a virtual replica of a physical system, process, or entity. It enables real-time monitoring, analysis, and optimization by simulating its physical counterpart [1]. This paradigm has found applications across diverse fields, including heat transfer systems [2], [3], healthcare [4], [5] solid mechanics [6], [7].

Despite these advancements, employing Digital Twins in the context of real-time thermal management—particularly under transient conditions—remains an ongoing challenge. This manuscript aims to advance the state of the art by introducing and rigorously evaluating two distinct, novel approaches for controlling the maximum temperature in a transient heat transfer system. Our key scientific contribution lies in the comparative analysis of these methods, offering new insights into their respective strengths, limitations, and computational efficiency. Specifically, we detail a Supervised Learning (SL) method, which leverages steady-state simulation data to infer a heat transfer coefficient for real-time thermal management, and a Reinforcement Learning (RL) strategy, where the algorithm learns on the fly through direct interaction with the unsteady Finite Element Method (FEM) model. By comparing these two methods, we provide insights into their respective capabilities and computational demands. The outline of the paper is the following. We first present the fundamental problem setting and boundary conditions for the transient thermal system, followed by a description of each digital twin approach. Subsequently, we compare and discuss the outcomes of both methods and then some conclusions are drawn.

2. Problem definition

The problem is governed by heat transfer equation in a solid medium, without any internal generation, in which a transient heat flux $q(t)$ is entering from the top boundary Γ_q. Lateral surfaces are adiabatic, while a heat transfer coefficient is applied on the bottom one Γ_c. So, the governing equations and related boundary conditions read as follows in eq. (4) and (5) respectively [8]:

$$\rho c \frac{\partial T}{\partial t} = \nabla \cdot (k \nabla T), \tag{1}$$

$$\begin{cases} -k\nabla T \cdot n = q(t), on\ \Gamma_q, \\ q_c(t) = h_c(t)(T - T_0), on\ \Gamma_c. \end{cases} \tag{2}$$

Heat flux time series is supposed to be known, while the heat transfer coefficient is unknown. It will be inferred by the DT-based on supervised learning or reinforcement learning approaches, as explained in sections 3 and 4. The computational domain and related boundary conditions are illustrated in Figure 1. In the absence of experimental data, the FEM model represents the physical system of the Digital Twin system, while the AI-based is its digital representation that has to control the physical twin. Second order finite elements and time discretization schemes are employed to numerically solve the present problem.

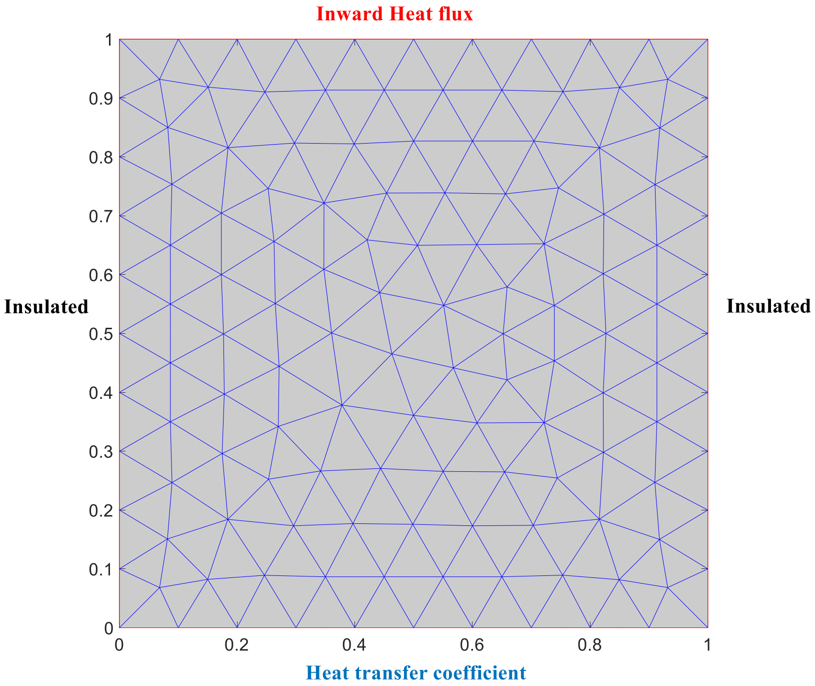

Figure 1. Computational domain and boundary conditions.

3. Supervised Learning approach

A Neural Network (NN) is used to the predict the heat transfer coefficient (label) from two inputs (features): the maximum temperature over the domain and the heat flux (see Figure 2). For reducing computational costs, the database that correlates the three quantities has been derived from steady-state simulations. In such an approach, three different steps can be identified:

- Dataset preparation (steady-state FEM simulations);
- Training, validation and testing of the Neural Network to correlate the maximum temperature, heat flux and heat transfer coefficient;
- The coupling between the transient FEM simulation and the NN and its test in an unseen (unsteady) scenario.

The dataset includes ten thousand results of numerical simulations in terms of random combination of heat transfer coefficient and heat flux boundary conditions. After dataset generation, the NN is trained and then is coupled to the unsteady FEM model. The coupling is made every 100 time-steps, and the heat transfer coefficient is constant with the time between two consecutive calls of the trained NN. The NN trained with steady-state data is used to control a transient thermal system, as steady-state data can establish an effective relationship between the heat transfer coefficient, heat flux, and maximum temperature. This approach significantly reduces computational costs in dataset generation and DT training. The key idea is that, even though the actual system is transient, the fundamental relationships between these physical quantities can be learned from steady-state data and applied to the transient system for controlling purposes. Furthermore, the heat transfer coefficient calculated by the DT is based on the maximum recorded heat flux during a time interval and will therefore be higher than that produced by a DT trained with transient data. This strategy, along with the use of a safety factor for the maximum temperature, enables effective temperature control in transient thermal systems, despite being trained on steady-state data. More details about the neural network architecture, the dataset, and the digital twinning structure can be found in ref. [9].

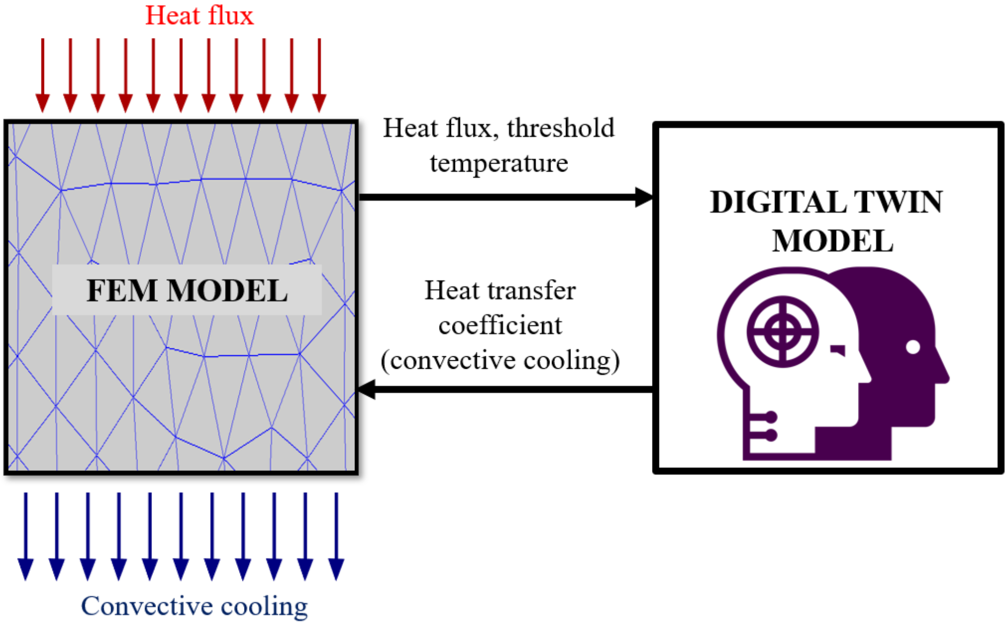
Heat flux
Heat flux, threshold temperature
FEM MODEL
DIGITAL TWIN MODEL
Heat transfer coefficient (convective cooling)
Convective cooling

Figure 2. DT based on supervised learning.

4. Reinforcement learning approach

Reinforcement Learning (RL) operates within the framework of Markov Decision Processes, where an agent interacts with an environment by taking actions based on observations and receives rewards as feedback. The environment's state evolves in response to the agent's actions, and the agent's goal is to learn a policy that maximizes cumulative rewards over time. In this context, the action is the heat transfer coefficient, while the observations are the instantaneous heat flux and the maximum temperature (the state of the environment). The flow-chart of a typical RL-based approach is presented in Figure 3.

Unlike the previous approach, AI and FEM are now tightly integrated and interconnected from the very beginning. Proximal Policy Optimization (PPO) is a popular RL algorithm that addresses the challenge of learning effective policies in complex environments. PPO aims to improve policy iteration by optimizing a surrogate objective function, which approximates the policy's performance and ensures stable training [10]. The RL agent's policy and value functions are approximated by a neural network with two hidden layers, each containing 64 neurons using rectified linear unit activations. The input layer processes a two-dimensional state consisting of the instantaneous heat flux and the domain's maximum temperature. A single continuous output action represents the heat transfer coefficient constrained to lie within the range [1, 15], thereby controlling the cooling process in order to keep maximum temperature within acceptable limits. An episode is defined to last a total 30 s with 900 timesteps per episode. At each step, the agent observes the current, selects an action and receives a reward based on how close the maximum temperature is to the target threshold. We train for a total of 10^6 timesteps, using a learning rate of 10^{-4}, a discount factor γ=0.99 a Generalized Advantage Estimator parameter λ=0.95. The clipped surrogate objective in PPO manages policy updates to stabilize training and promote efficient exploration. The reward function plays a crucial role in guiding the agent's behaviour towards achieving the desired objectives. By appropriately designing the reward function, such as penalizing deviations from desired temperature thresholds and rewarding actions that lead to effective heat transfer coefficients, the agent can learn to optimize the heat transfer coefficient and control the maximum temperature effectively. In practical terms, we consider three different scenarios for the proposed transient thermal system. In the first case, we apply a positive reward equal to 1 in case the distance between the maximum temperature and the threshold value is less than one (equation 3a). In the second case, if the maximum temperature overcomes the limit value, a negative reward (penalty) is given, proportional to their difference. However, the penalty is mitigated by the heat transfer coefficient to inform the agent that higher heat transfer coefficients have a beneficial effect in reducing the maximum temperature (equation 3b). Finally, if the maximum temperature is less than the threshold value (but more than 1), we assign a positive reward equal to 1, with an attenuation term which increases with the increase in the heat transfer coefficient (equation 3c). This is to inform the agent that a lower heat transfer coefficient may be explored to optimize the cooling effect. The calibration/scaling terms in the equation (3) are introduced in the user defined reward function by trial and error to take into account the different range of variations for the temperature and heat transfer. The value of the coefficient of the logarithmic function including the heat transfer coefficient in equation (3a) is higher the coefficient in (3b) because a more importance is given to stay below the threshold value maximum temperature. A similar reward function (without taking into consideration the heat transfer coefficient) for the optimization of thermal management of a battery can be found in ref. [11].

$$reward = \begin{cases} 1\ (a), \\ -\,|T_{max} - T_{threshold}| \cdot 0.06 + 0.07 \cdot \log(h_c + 1)\ \ (b), \\ 1 - 0.01 \cdot |T_{max} - T_{threshold}| - 0.05 \cdot \log(h_c + 1)\ (c). \end{cases} \tag{3}$$

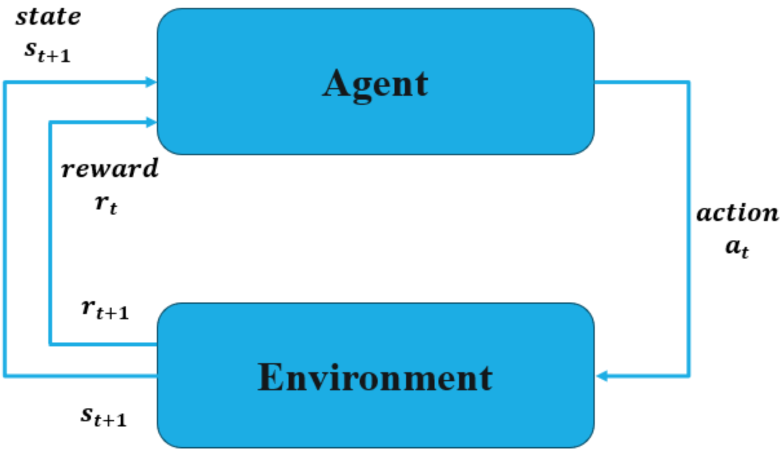

Figure 3. DT based on reinforcement learning.

5. Results and discussion

To make a fair comparison between the two approaches, the same transient heat flux as boundary condition is adopted. A baseline sinusoidal distribution is chosen with some random noise on both amplitude and offset of the function, as follows:

$$q(t) = A_0(t) + A_1(\mathrm{t})\left[\sin\left(\frac{2\pi f}{30}t\right) + 1\right], \tag{4}$$

where A_0, A_1 vary randomly between [0, 70] and [0,30] respectively at every time step and $f = 1/30$. Such a function allows the agent to have a complete picture of the possibles trends of the heat flux in the 0-100 range.

In the supervised learning approach, after the dataset is built and the NN is successfully trained, validated and tested, a correlation between heat transfer coefficient, maximum temperature and heat flux is available, as shown in Figure 4 which presents the FEM results in terms of the maximum temperature versus the heat transfer coefficient for three different heat fluxes. Every 9 time steps, this correlation is invoked to calculate the heat transfer coefficient that is needed to solve the transient FEM problem. The inputs of the NN are the maximum heat flux over a previous interval and the threshold temperature, fixed at 100. The results of this approach are shown in Figure 5. It can be noted that maximum temperature is always drastically lower than the fixed threshold value. The reason is behind the conservative choice of using a steady-state approach. It does not consider any delay effects in the temperature variation with the heat flux, that may lead to lower heat transfer coefficient.

On the other hand, in the reinforcement learning-based control strategy, this phenomenon is intrinsically modelled as the training of the agent is made on the fly, on the purely transient data. The learning stopped once the cumulative reward function, described by the equation (3), has reached a plateau, as shown in Figure 6. In Figure 7, the trend of maximum temperature and heat transfer coefficient are plotted. It is clear that the thermal system reaches higher maximum temperatures compared to the previous approach, but always lower than the maximum allowed value. As a consequence, heat transfer coefficient values are significantly lower than the previous case. The drawback is the higher computational costs of RL-based approach, given the coupled nature of the problem. Figure 8 compares the two maximum temperatures during the time for both approaches.

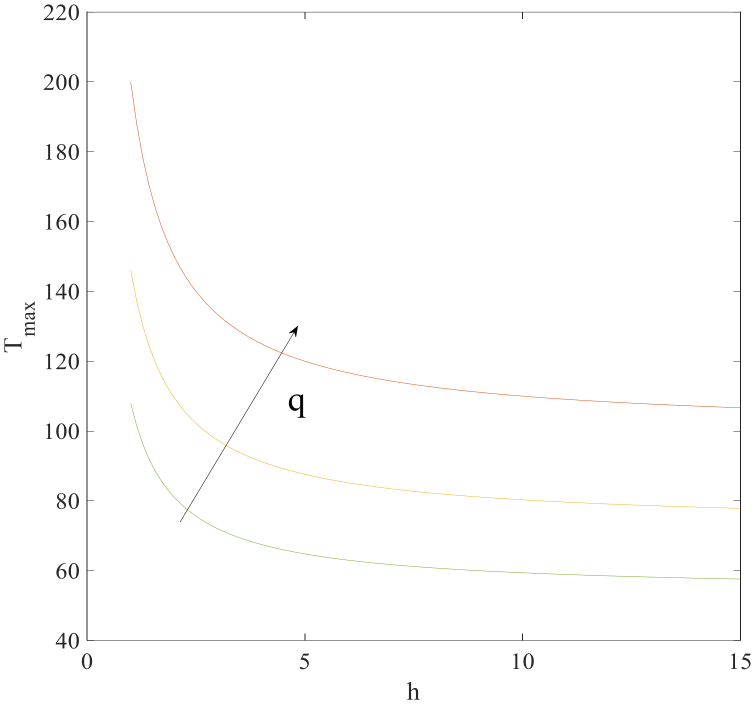

Figure 4. Supervised learning approach: correlation between T_{max}, q, h_c.

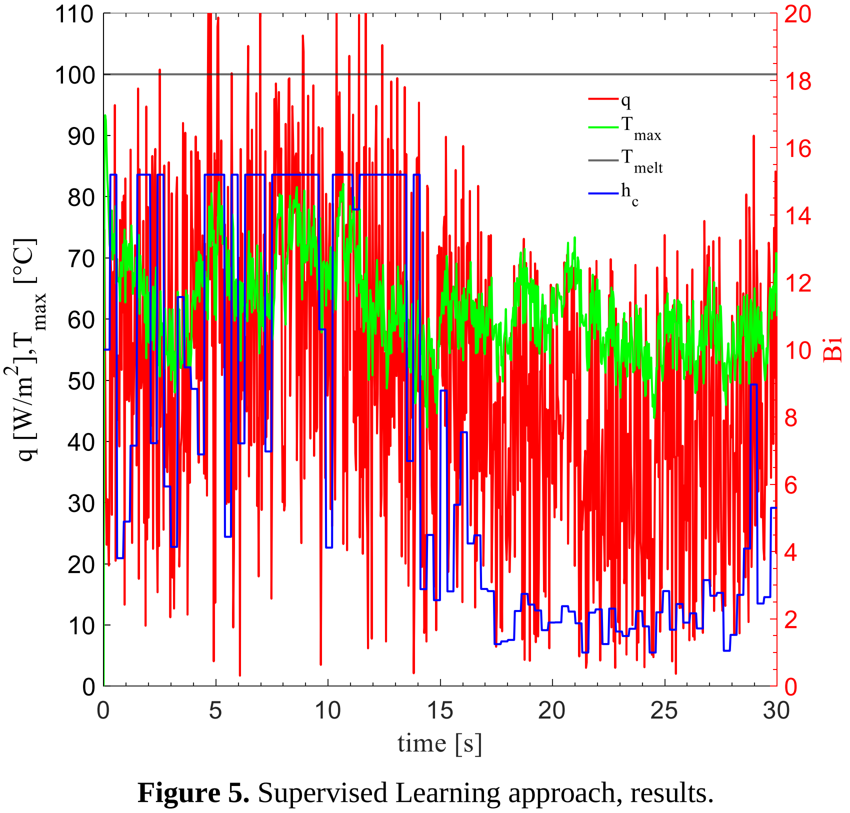

Figure 5. Supervised Learning approach, results.

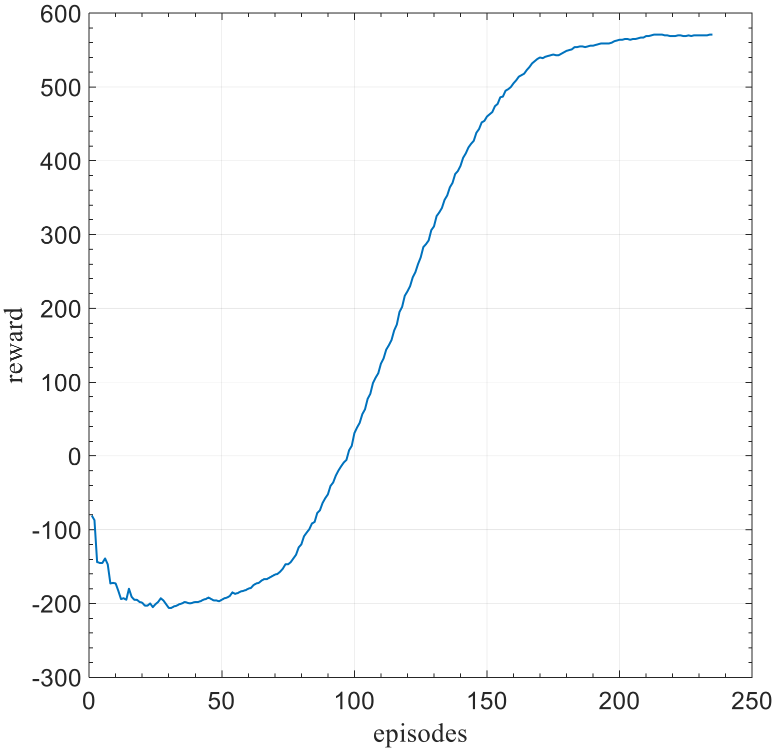

Figure 6. Reinforcement learning approach: cumulative mean reward function.

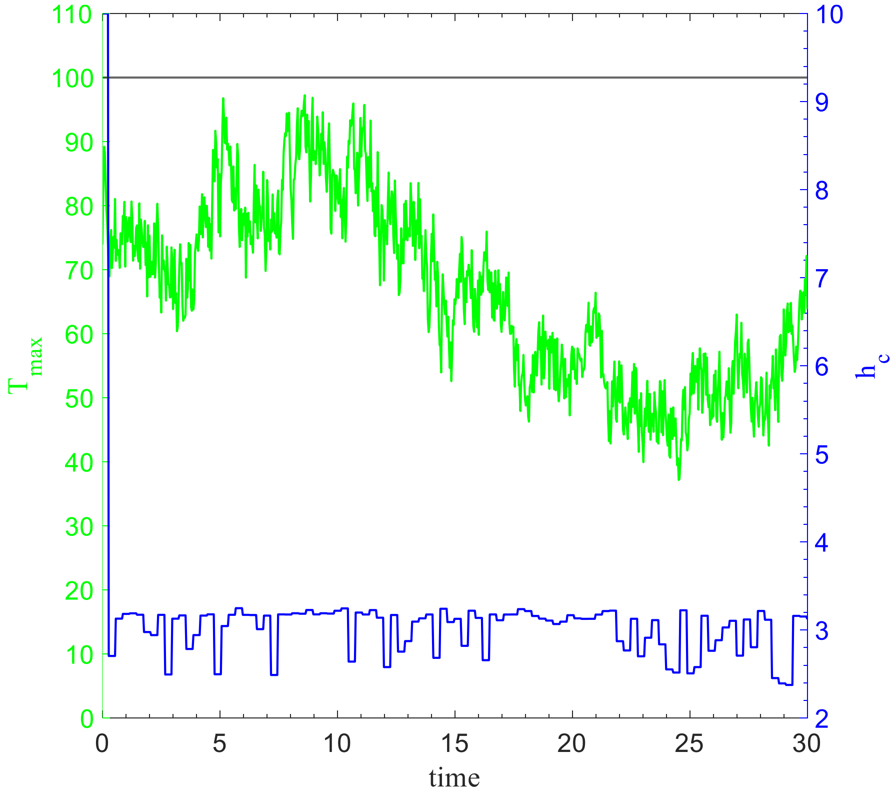

Figure 7. Reinforcement learning approach, results.

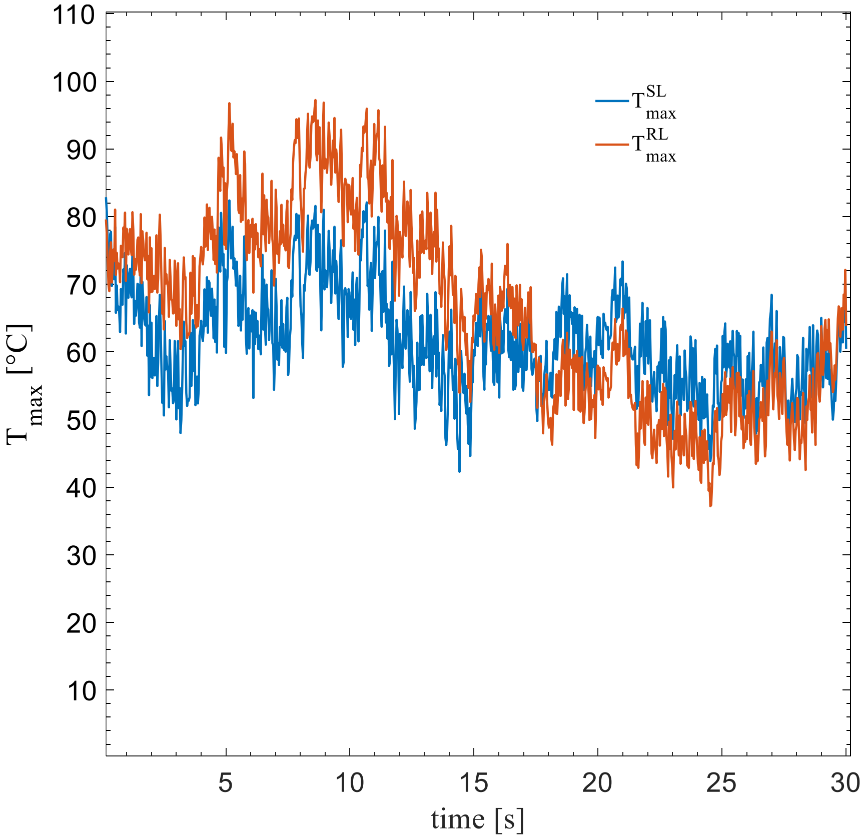

Figure 8. Comparison between RL and SL in terms of maximum temperature.

6. Conclusions

In this article, two approaches of digital twinning for transient thermal system are proposed. The problem to which the proposed approaches are applied is a 2D heat transfer problem, where the inward transient heat flux is known. The heat transfer coefficient has to be determined to allow a maximum temperature of the system below a threshold fixed value. The first one, based on Supervised Learning requires a dataset generation from steady-state FEM simulations. Then, a neural network is trained to correlate heat transfer coefficient, maximum temperature and heat flux. The third step regards the effective coupling between FEM simulation (the physical twin) and its digital replica to effectively control the physical system. On the other hand, the Reinforcement Learning approach requires the strong interaction between the environment (the FEM-based thermal simulation) and the action made by the agent (the heat transfer coefficient). The results, obtained with the same noisy sinusoidal heat flux boundary condition, show that in both cases the control is effectively working. In the first case, the maximum temperature is quite farther from the threshold value compared to the second approach.

7. References

[1] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital Twin: Enabling Technologies, Challenges and Open Research," *IEEE Access*, vol. 8, pp. 108952–108971, 2020, doi: 10.1109/ACCESS.2020.2998358.

[2] W. Bielajewa, M. Tindall, and P. Nithiarasu, "COMPARATIVE STUDY OF TRANSFORMER- AND LSTM-BASED MACHINE LEARNING METHODS FOR TRANSIENT THERMAL

FIELD RECONSTRUCTION," *Computational Thermal Sciences: An International Journal*, vol. 16, no. 3, 2024.

[3] H. R. Tamaddon-Jahromi, N. K. Chakshu, I. Sazonov, L. M. Evans, H. Thomas, and P. Nithiarasu, "Data-driven inverse modelling through neural network (deep learning) and computational heat transfer," *Comput Methods Appl Mech Eng*, vol. 369, 2020, doi: 10.1016/j.cma.2020.113217.

[4] N. K. Chakshu, I. Sazonov, and P. Nithiarasu, "Towards enabling a cardiovascular digital twin for human systemic circulation using inverse analysis," *Biomech Model Mechanobiol*, vol. 20, no. 2, pp. 449–465, 2021, doi: 10.1007/s10237-020-01393-6.

[5] N. K. Chakshu and P. Nithiarasu, "An AI based digital-twin for prioritising pneumonia patient treatment," *Proc Inst Mech Eng H*, vol. 236, no. 11, pp. 1662–1674, 2022, doi: 10.1177/09544119221123431.

[6] S. Vlase, M. Marin, M. L. Scutaru, and R. Munteanu, "Coupled transverse and torsional vibrations in a mechanical system with two identical beams," *AIP Adv*, vol. 7, no. 6, Jun. 2017, doi: 10.1063/1.4985271.

[7] M. Marin, A. Öchsner, and M. M. Bhatti, "Some results in Moore-Gibson-Thompson thermoelasticity of dipolar bodies," *ZAMM Zeitschrift fur Angewandte Mathematik und Mechanik*, vol. 100, no. 12, Dec. 2020, doi: 10.1002/zamm.202000090.

[8] A. Bejan and A. D. . Kraus, *Heat transfer handbook*. John Wiley, 2003.

[9] A. Di Meglio, N. Massarotti, and P. Nithiarasu, "A physics-driven and machine learning-based digital twinning approach to transient thermal systems," *Int J Numer Methods Heat Fluid Flow*, 2024, doi: 10.1108/HFF-10-2023-0616.

[10] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," Jul. 2017, [Online]. Available: http://arxiv.org/abs/1707.06347

[11] L. He *et al.*, "Optimization of thermal management performance of direct-cooled power battery based on backpropagation neural network and deep reinforcement learning," *Appl Therm Eng*, vol. 258, p. 124661, Jan. 2025, doi: 10.1016/J.APPLTHERMALENG.2024.124661.