

ML-based surrogate cure simulation for predicting process time and temperature overshoot in Resin Transfer Moulding

Evrydiki Kyriazi^{*1}, Georgios Petsinis², Charalampos Zervos^{1,2}, Ioannis Pouloupoulos¹, Georgios Syriopoulos¹, Geoffrey Neale³, Mehdi Asareh³, Swaroop Narayanan⁴, Christopher Page⁴, Stuart Lewis⁴,
Hercules Avramopoulos¹, Alexandros Skordos³

¹School of Electrical & Computer Engineering, National Technical University of Athens

² Talos Analytics IKE, Athens, Greece

³Faculty of Engineering and Applied Sciences, Cranfield University, Cranfield, Bedford MK43 0AL, UK

⁴ TWI, Cambridge CB21 6AL, UK

*Corresponding author's email: evkyriazi@mail.ntua.gr

Abstract

The cure stage of thermosetting composites production is critical for the overall process duration and manufacturing costs. While process simulations are commonly used to estimate cure behaviour, real-time predictive capabilities in Resin Transfer Moulding (RTM) remain limited, primarily due to the computational cost of finite element (FE) methods. To address this gap and accurately estimate cure process parameters in RTM, this study proposes a surrogate cure simulation approach based on two state-of-the-art machine learning (ML) voting ensemble models—XGBoost and Light Gradient Boosting Machine—designed to predict cure time and temperature overshoot. To train the model, the cure of an epoxy/carbon fibre flat plate was simulated using the FE solver Marc, providing data over a wide range of conditions. The predictions of temperature overshoot and cure time demonstrate remarkable consistency and high accuracy (R^2 values up to 98 %) with execution times under 30 ms for both variables. Performance was validated against unseen simulation data and further verified through RTM manufacturing trials and differential scanning calorimetry (DSC), confirming cure completion at a glass transition temperature of 191–194 °C. Unlike existing studies that remain simulation-focused, this approach bridges process simulation and data-driven modelling, offering a practical tool for real-time optimisation in industrial RTM applications.

Keywords: A. Machine Learning, B. Polymer-matrix composites (PMCs), C. Process Simulation, D.

Cure

1. Introduction

Continuous fibre thermosetting matrix composites play a major role in the engineering of advanced structures due to their favourable combination of light weight and high strength and stiffness, as well as the design flexibility and opportunities offered by their inherent anisotropy. The aerospace industry leads the way in applications of advanced composites with other sectors, including wind energy and automotive, presenting significant growth in recent years. The high manufacturing cost and the inherently low production rates are the main limiting factors in increasing adoption of composites. Current research focuses on rate intensification through the development of manufacturing strategies, materials and methodologies of process optimisation and control.

The cure stage of thermosetting composites production is a major contributor to both process duration and manufacturing costs. Curing is a – usually thermally activated – process that transforms the polymeric matrix from a low-viscosity liquid to a glassy solid through a crosslinking reaction. Its duration is governed by the chemical kinetics of the matrix crosslinking reaction and by heat transfer effects occurring in the manufacturing assembly. The outcome is critical for component quality as it governs the final glass transition temperature of the matrix, the generation of residual stress and the occurrence of distortion. The exothermic character of the curing reaction presents a significant challenge. In combination with the low through thickness thermal conductivity of composites, it can lead to temperature overshoots which are intensified with increasing component thickness, potentially leading to a lack of uniformity and challenges in process control. In extreme cases, this can lead to detrimental matrix degradation. To overcome these challenges, the composites industry adopts manufacturer recommended cure cycles (MRCCs) which, in order to achieve a robust process and to produce components that satisfy stringent certification requirements, are developed using conservative assumptions and large safety margins. This leads to long processing cycles. Increasing the confidence in designing and controlling the process through efficient and accurate process simulation creates an opportunity for optimisation to increase the overall efficiency of production.

Numerical optimisation has been used successfully to reduce process duration by optimising the cure cycle. A variety of single-objective optimisation methods ranging from gradient-based minimisation [1] to zero order search [2,3] and Genetic Algorithms (GAs) [4,5] have demonstrated benefits in process duration that can reach a reduction of up to 60 %. Extension of the optimisation problem to a multi-objective

setting addressing simultaneously the minimisation of process duration and temperature overshoot or cure state uniformity [6,7] has shown that it is possible to achieve significant benefits with respect to both objectives, delivering improvements in both efficiency and product quality.

Numerical cure optimisation methodologies operate based on the iterative solution of an FE model of cure, which is a computationally expensive process. Furthermore, cure simulation is an idealised representation of the actual process in which aspects of uncertainty due to material or process conditions variability are not taken into account. Incorporation of variability is possible through stochastic simulation [8] and integration of data from on-line sensing [4] in the optimisation loop. However, both of these further inflate the computational complexity of the problem. Therefore, a combination of process optimisation with real-time information is extremely challenging to achieve. This makes the development of fast and precise surrogate models a critical development.

Computationally efficient representations of the cure problem, also offering flexibility in the mapping of inputs and outputs of the problem, are necessary to enable utilisation of optimisation methods in the context of real-time control of curing. The demonstrated benefits of using approaches based on physics-driven numerical modelling solutions such as FE analysis are subject to limitations due to execution times and computation resources required to perform cure simulation. The non-linear, coupled and time-dependent nature of the problem makes it highly demanding computationally, necessitating the use of approximating assumptions to enable iterative function evaluation, especially when envelope predictions are required. Therefore, the development of ML techniques has recently received attention, enabling the reduction of computation time and enhancing the simulation performance for polymer composite parts [9]. The accuracy of ML models is linked to the need for a large amount of data for better statistical analysis and generalisation. As an alternative to experimental data, which are practically limited in terms of population size, synthetic data generated by numerical simulation based on FE modelling can be used for this purpose [10].

Given the limitations of real-time information from sensor measurements, data assimilation exploiting neural networks (NN) has been used to perform cure optimisation, carrying out an internal estimation of thermal conductivity to perform a more accurate simulation [7]. In particular, machine learning has been frequently employed to predict various mechanical properties and characteristics of composite materials. Flow Front Net – a deep learning approach – has been utilised to capture spatial

disturbances in RTM to predict dry spots with an accuracy of 91.7 % [11]. Extreme Gradient Boosting and Light Gradient Boosting Machine, which are gradient boosting ML algorithms that minimise a loss function by iteratively adding weak learners (decision trees) to the ensemble tree, are the most accurate models, achieving 85 % and 83 % accuracy, respectively, in predicting the filling quality of RTM using synthetic data [10]. The incorporation of physics-based features in the training of NN models introduces a bias towards known domain-specific behaviour [12,13]. Theory-guided ML (TGML) models have been used to predict the characteristics of the bond-line of honeycomb sandwich structures with a 5056-point dataset based on decision tree-based regressors to optimise the cure profile used in autoclave co-curing [12,13]. A comparison between a classic NN algorithm and a TGML model shows that incorporation of theory-based bias improves accuracy and efficiency of the algorithm [12,13]. Among more recent ML developments, an image-based Pixel RNN surrogate for RTM flow trained on FE simulation data, has been demonstrated to reach up to 97% accuracy [14]. An ANN-based method reported in [15] achieved a mean error of $\sim 1.6\%$ in predicting in-situ thermal profiles, whilst the work reported in [16] trained models on approximately 2,000 FE-simulated curing histories (varying temperature ramps, hold times) to map cure-cycle inputs to degree-of-cure vs. time trajectories. A deep-CNN surrogate model has also been used for fast prediction of cure-induced residual stress in polymer-matrix composites [17]. Finally, a physics-informed ML framework for composite curing that co-trains multiple neural networks has been proposed [18], improving model interpretability and generalisation.

The aim of this study is to improve the curing process by utilising the process objective determined with ML techniques. This approach aims to replace the time-consuming FE methods potentially simplifying and enhancing efficiency in optimisation. Previous research has demonstrated that there are significant benefits that can be achieved by optimising thermal profiles, and that online application of optimisation reflects the improvement and efficiency opportunities found using off-line analysis. However, the component geometries that can be addressed with a full FE representation of the cure process are limited to those compatible with a 1-D model that can be executed within seconds [19]. The present study focuses on the development and validation of two individual voting ensemble models incorporating the Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) to predict the temperature overshoot and cure time of various cure profiles in RTM. The cure of an epoxy/carbon fibre flat plate has been simulated in order to generate the datasets fed into the ML process. The ML models were

selected and trained using the FE simulation data. Following the training phase, the proposed ML models underwent further testing and optimisation to determine their predictive accuracy for unseen data. The overall development was validated using experimental results from process trials and DSC assessment of final glass transition temperature. Overall, the contribution of this work lies in the relatively limited literature on RTM curing, presenting a modular, time-efficient ML-based approach for process modelling. The method reports statistical confidence in predictions and is supported by two thin-laminate thickness trials, offering a proof-of-concept with scope for extension to other thicknesses and geometries, and representing a potential capability for reliable process design in composite manufacturing.

2. Methodology

2.1. Cure simulation

The materials used in this study are Hexcel HexFlow® RTM6-2 epoxy resin and HTA40 carbon fibre. The material models necessary for the cure simulation cover chemical and thermal behaviour and include cure kinetics and specific heat capacity and thermal conductivity and their dependence on degree of cure and temperature. The cure kinetics model is a central component governing the accuracy of the simulation. For some of the properties the dependence on temperature and degree of cure is governed by material state, i.e., whether the material is under or over its instantaneous glass transition temperature. Therefore, a model linking glass transition temperature to degree of cure is also necessary.

The cure kinetics model for the Hexcel HexFlow® RTM6 resin system is a combination of two mechanisms expressed by an nth term and an autocatalytic term [20]:

$$\frac{d\alpha}{dt} = k_1(1 - \alpha)^{n_1} + k_2(1 - \alpha)^{n_2}\alpha^m \quad (1)$$

Here, α is the degree of cure; m is the autocatalytic reaction order, n_1 and n_2 are the reaction orders of nth order term and autocatalytic term, respectively. Also, k_1 and k_2 are the reaction rate constants. These are defined as follows:

$$\frac{1}{k_i} = \frac{1}{k_{i,C}} + \frac{1}{k_d}, \quad i = 1, 2 \quad (2)$$

where $k_{i,C}$ are Arrhenius functions of temperature for the chemical reaction and k_d is the rate constant of diffusion. The rate constants are expressed as follows:

$$k_{i,C} = A_i e^{(-E_i/RT)}, \quad i = 1, 2 \quad (3)$$

$$k_d = A_d e^{(-E_d/RT)} e^{(-b/f)} \quad (4)$$

Here A_1 , A_d are pre-exponential factors, where A_1 the pre-exponential factor of the nth order term, A_2 of the autocatalytic term and A_d corresponding to diffusion and b is a fitting parameter. E_1 and E_d are the activation energies for the chemical reactions ($E_1 = 72,900 \text{ Jmol}^{-1}$ and $E_2 = 57,820 \text{ Jmol}^{-1}$) and for diffusion respectively. T is the absolute temperature, R the universal gas constant and f the equilibrium free volume, which is expressed as follows:

$$f = w(T - T_g) + g \quad (5)$$

Here w is the slope and g is intercept constant of the equilibrium free volume model, and T_g is the instantaneous glass transition temperature following the Di Benedetto equation [21]:

$$T_g = T_{g0} + \frac{(T_{g\infty} - T_{g0})\lambda\alpha}{1 - (1 - \lambda)\alpha} \quad (6)$$

Here $T_{g\infty}$ and T_{g0} are the glass transition temperature of the fully cured and uncured material and λ is a parameter controlling the convexity of the glass transition versus degree of cure curve. Constants of the models represented by Eqs. (1)-(6) are reported in Table 1 [22].

Table 1: Parameter values for the RTM6 resin cure kinetics and glass transition temperature development models [22].

Parameter		Value
Pre-exponential factor of the nth order term	A_1	$19,000 \text{ s}^{-1}$
Pre-exponential factor of the autocatalytic term	A_2	$22,080 \text{ s}^{-1}$
Pre-exponential factor of diffusion	A_d	$6.76 \cdot 10^{18} \text{ s}^{-1}$
Activation energy of the nth order term	E_1	$72,900 \text{ Jmol}^{-1}$
Activation energy of the autocatalytic term	E_2	$57,820 \text{ Jmol}^{-1}$
Activation energy of diffusion	E_d	$138,000 \text{ Jmol}^{-1}$
Autocatalytic reaction order	m	1.29
Reaction order of the nth order term	n_1	1.97
Reaction order of the autocatalytic term	n_2	1.53
Exponent of diffusion term	b	0.452
Equilibrium free volume model slope	w	0.00048 K^{-1}
Equilibrium free volume model intercept	g	0.025
Glass transition temperature of uncured material	T_{g0}	$-11 \text{ }^\circ\text{C}$
Glass transition temperature of fully cured material	$T_{g\infty}$	$206 \text{ }^\circ\text{C}$
Glass transition temperature convexity constant	λ	$0.435 \text{ Jg}^{-1}\text{ }^\circ\text{C}^{-2}$

The specific heat capacity of the composite is calculated using the rule of mixtures:

$$c_p = w_f c_{pf} + (1 - w_f) c_{pr} \quad (7)$$

where w_f is the fibre weight fraction, c_{pf} the fibre specific heat capacity and c_{pr} the specific heat capacity of the resin. The specific heat capacity of the resin and the fibre are computed using [23]:

$$c_{pf} = A_{fc_p} T + B_{fc_p} \quad (8)$$

$$c_{pr} = A_{rc_p} T + B_{rc_p} + \frac{\Delta_{rc_p}}{1 + e^{C_{rc_p}(T - T_g - \sigma)}} \quad (9)$$

where A_{fc_p} and B_{fc_p} control the linear dependence of fibre specific heat capacity on temperature, A_{rc_p} and B_{rc_p} represent the linear dependence of specific heat capacity of the epoxy on temperature and Δ_{rc_p} , C_{rc_p} and σ are the strength, width and temperature shift of the specific heat capacity step occurring at vitrification. Also, T is the absolute temperature and T_g is the instantaneous glass transition temperature, as defined in Eq. (6). The values of the parameters involved for the resin system of this study in Eqs. (7)-(9) are reported in Table 2 [23].

Table 2: Model parameter values for the specific heat capacity [23], thermal conductivity [4], and constitutive models.

Parameter		Value
Carbon fibre specific heat capacity model slope	A_{fc_p}	$0.0023 \text{ Jg}^{-1}\text{°C}^{-2}$
Carbon fibre specific heat capacity model intercept	B_{fc_p}	$0.765 \text{ Jg}^{-1}\text{°C}^{-2}$
Resin specific heat capacity model slope	A_{rc_p}	$0.0025 \text{ Jg}^{-1}\text{°C}^{-2}$
Resin specific heat capacity model intercept	B_{rc_p}	$1.8 \text{ Jg}^{-1}\text{°C}^{-2}$
Resin specific heat capacity model step	Δ_{rc_p}	$-0.25 \text{ Jg}^{-1}\text{°C}^{-2}$
Resin specific heat capacity model step breadth parameter	C_{rc_p}	1.1 °C^{-1}
Resin specific heat capacity model step shift parameter	σ	16.5 °C
Carbon fibre transverse thermal conductivity	B_{tf}	$0.84 \text{ Wm}^{-1}\text{°C}^{-2}$
Resin thermal conductivity model quadratic coupling	a_{Kr}	$0.0008 \text{ Wm}^{-1}\text{°C}^{-2}$
Resin thermal conductivity model coupling constant	b_{Kr}	$-0.0011 \text{ Wm}^{-1}\text{°C}^{-2}$
Resin thermal conductivity model linear temperature constant	c_{Kr}	$-0.0002 \text{ Wm}^{-1}\text{°C}^{-2}$
Resin thermal conductivity model quadratic conversion constant	d_{Kr}	$-0.0937 \text{ Wm}^{-1}\text{°C}^{-2}$
Resin thermal conductivity model linear conversion constant	e_{Kr}	$0.22 \text{ Wm}^{-1}\text{°C}^{-2}$
Resin thermal conductivity model intercept	f_{Kr}	$0.12 \text{ Wm}^{-1}\text{°C}^{-2}$

The thermal conductivity of the composite in the through thickness direction is computed using a homogenisation model [24]:

$$K_{33} = v_f K_r \left(\frac{K_{tf}}{K_r} - 1 \right) + K_r \left(\frac{1}{2} - \frac{K_{tf}}{2K_r} \right) + K_r \left(\frac{K_{tf}}{K_r} - 1 \right) \sqrt{v_f^2 - v_f + \frac{\left(\frac{K_{tf}}{K_r} + 1 \right)^2}{\left(\frac{2K_{tf}}{K_r} - 2 \right)^2}} \quad (10)$$

where K_r is the resin thermal conductivity, K_{tf} is the thermal conductivity of the fibre in the transverse direction, and v_f is the fibre volume fraction. The thermal conductivity of the carbon fibre in the transverse direction is:

$$K_{tf} = B_{tf} \quad (11)$$

where B_{tf} is a constant. The thermal conductivity for the RTM6 epoxy resin system is a function of degree of cure and temperature and is expressed as [4]:

$$K_r = a_{Kr} T \alpha^2 + b_{Kr} T \alpha + c_{Kr} T + d_{Kr} \alpha^2 + e_{Kr} \alpha + f_{Kr} \quad (12)$$

Here a_{Kr} , b_{Kr} , c_{Kr} , d_{Kr} , e_{Kr} , f_{Kr} are coefficients of the polynomial function. The parameters of the thermal conductivity model are reported in Table 2 [4].

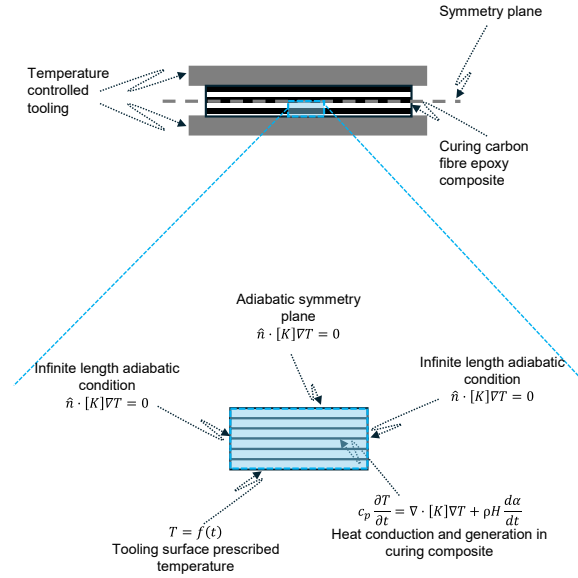


Figure 1: Cure assembly and FE model representation.

A heat transfer simulation model was implemented in the FE solver Marc 2023.2 by Hexagon [25] to simulate the cure of an RTM6/carbon flat plate of a thickness of 12 mm, cured with double-sided heating.

The model setup was carried out using the Marc Mentat 2023.2 pre-processor. The Marc solver allows the solution of the coupled thermo-chemical problem of cure, also allowing the incorporation of a degree of cure material state variable, with its evolution directly linked to a cure kinetics model. The fibre volume fraction of the composite is 60 %, corresponding to a density of 1.5 g/cm³, and the total heat of reaction of the resin 465 J/g. Figure 1 shows a schematic of the setup and the corresponding model representation. The FE model domain corresponds to one half of the composite part due to symmetry. The large in-plane dimensions compared to the thickness of the part with aspect ratios exceeding 20 allow use of an adiabatic condition on the side boundaries of the domain. A perfect insulation boundary condition applied at the mid-plane ensures symmetry of the temperature distribution. The domain is represented by 6 3D iso-parametric eight-noded composite brick elements (175 element type in Marc [26]). Each element represents one layer of the composite with 1 mm nominal thickness. The 3D elements are only active in the thickness direction, as the in-plane symmetry of the problem implies uniform temperature at each layer of nodes. A time dependent prescribed temperature boundary condition is applied to the lower boundary of the curing material to represent a two-dwell cure profile. The prescribed profile comprises a heating ramp from an initial temperature of 120°C to the first isothermal temperature, the first isothermal segment, a ramp to the second isothermal temperature and the second isothermal segment. The boundary condition representing the cure profile was implemented using the user subroutine FORCDT, which allows definition of temperature as a function of time [27]. FORCDT is executed by the solver at each increment for each node that is part of the prescribed temperature boundary. User subroutines UCURE, USPCHT, and ANKOND [27] were used for the cure kinetics, specific heat capacity and thermal conductivity material sub-models expressed by Eqs. (1-12). These use the degree of cure as an underlying state variable, which is available through the CURING option of Marc. Alongside temperature, these provide the independent variables of the functional dependence of reaction rate, thermal conductivity and specific heat capacity. UCURE, USPCHT, and ANKOND are called in the element loop for each integration point at each increment of the solution. The interaction of these subroutines with the central solver and its element and time loops is represented schematically in Figure 2.

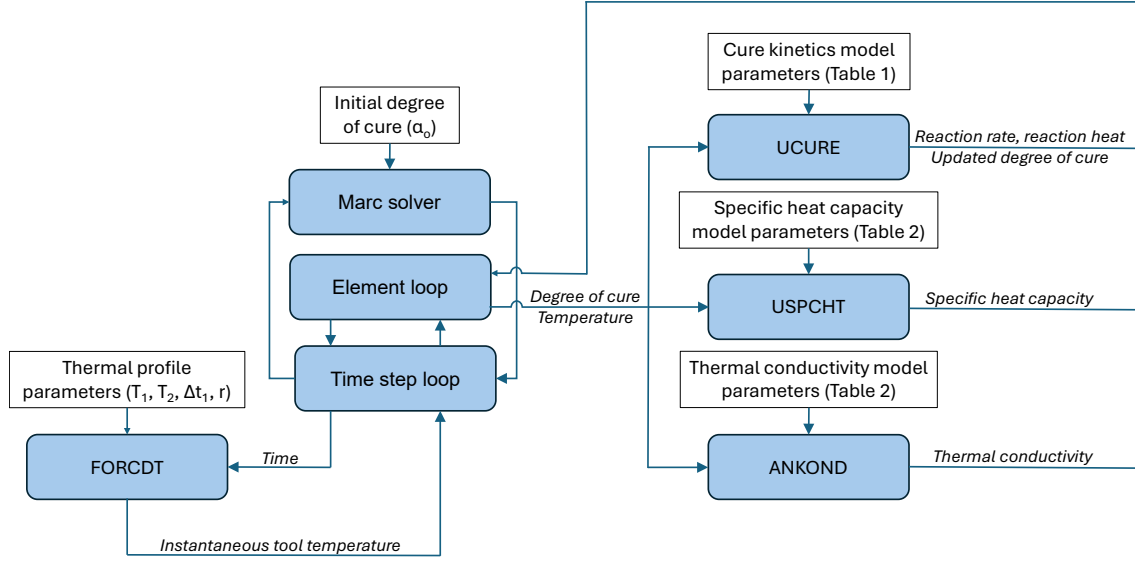


Figure 2: Flowchart of the interaction between the user subroutines and solver of the FE model.

2.2. Dataset generation

The simulation described in the previous section was utilised to generate three datasets, one for the training stage and the other two for validation, mapping the dependence of process outcomes on the cure cycle followed. The thermal profile was parameterised using four parameters (Figure 3):

- i. Temperature of the first isothermal segment (T_1);
- ii. Temperature of the second isothermal segment (T_2);
- iii. Duration of the first isothermal segment (Δt_1);
- iv. Ramp rate (r).

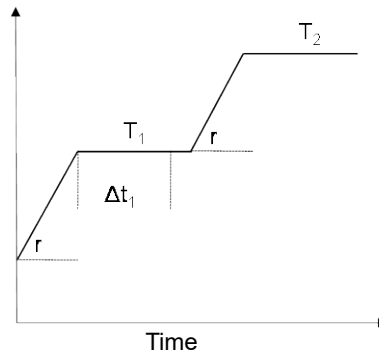


Figure 3: Parametrisation of the cure profile.

The duration of the second isothermal segment is not considered as a design parameter in the analysis due to the definition of cure process duration as the time at which the minimum degree of cure

throughout the curing component reaches a threshold set at 90 % for the resin system of this study. When this threshold is reached throughout the component, the cure is considered complete, and the simulation of the process is interrupted. A 90 % threshold has been chosen as this is approximately the maximum degree of cure that the epoxy resin of this study reaches during its MRCC [28], [29] according to the cure kinetics model used in this study. The process parameters considered here are the primary drivers of the cure process in double-sided closed mould RTM. Other variables involved in the overall RTM process, such as injection pressure vacuum, resin viscosity, fabric permeability and injection temperature do not affect the cure stage, which is initiated once filling is complete and is governed by the thermal profile and the cure and thermal properties of the material. Furthermore, the double-sided nature of the RTM process means that interaction with the environment in terms of heat transfer and the corresponding parameters, such as air temperature and surface heat transfer coefficient of the outer mould surfaces, do not influence the thermal field of the curing composite.

Material uncertainty is incorporated in the analysis by considering a range of values for the initial degree of cure of the material at the beginning of the cure process, which is included as a fifth input parameter, alongside the four parameters defining the thermal profile. This follows from the outcomes of variability analysis of the cure kinetics of the RTM6 epoxy resin system, which have established that the initial cure state is the main factor of uncertainty [22]. Other cure process variability is incorporated directly through the consideration of tool temperature, which alongside the surface heat transfer coefficient of tooling are the most influential parameters with uncertainty [30]. The latter is of no significance here, due to the double-sided nature of the RTM process. Two process metrics (Cure completion time and Maximum temperature overshoot) have been obtained by the simulation for each point in the dataset to be used in training and validation. The importance of cure duration is directly linked to efficiency and process cost. An increased duration of cure is translated to longer use of capital equipment, increased energy consumption to maintain a high temperature in the cure assembly and potentially increased staff time for monitoring the process. The maximum overshoot is important in terms of part quality. A greater overshoot generates a greater thermal gradient that drives thermal expansion and contraction mismatch and residual stress. Therefore, as overshoot increases the risk of matrix microcracking due process stress also increases, adversely affecting both product mechanical performance and environmental resistance. In extreme scenarios, in which the overshoot results in temperatures exceeding the matrix degradation threshold,

scission of cross-linking bonds can occur causing a reduction of glass transition temperature and an associated knockdown of high temperature properties of the material as well as lower environmental resistance. The occurrence of extremely high exothermic overshoots combined with the self-accelerating nature of the curing reaction can also lead to major safety incidents associated with uncontrollable resin degradation.

The datasets generated are summarised in Table 3. These include a training dataset – the entire data in the phase of training phase – with 32,768 points created for both the training and testing stage, as well as two validation sets. The testing set is 15 % of the entire dataset and is used during the model training phase but is not part of the actual training data. Its role is to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The training dataset contains the values of seven parameters (four thermal profile parameters, initial degree of cure and two simulation outcomes). The input variables of the datasets were generated using random vectors, over the specified range for each parameter, as specified in Table 4. The selection of dwell temperatures is based on the recommended cure cycles temperature of 160 °C and 180 °C for the resin system [29,31]. Similarly, the heating rate is using the recommended profile heating rate of 1 °C/min [29], as a minimum, with a potential increased up to 4 °C/min for enhanced efficiency. The first dwell duration range can exceed the nominal first dwell duration of 75 min [29] by 50 min with a lower bound of 0 min to represent the case of a single dwell. The initial degree of cure range represents a swing of 3 standard deviations in the initial state, which has been estimated to be 0.6 % [22], over a nominal value of 3 %.

Table 3:Size of provided datasets.

Dataset	Number of points	Purpose of data
Training set	32768	Entire dataset used in the training phase to train and test the model.
Regular set	1200	Testing of the performance within the training feature space
Random extreme set	1200	Testing of the robustness outside of the training feature space

Table 4: Training data range; $U(0,1)$ denotes a uniformly distributed random variable in $[0,1]$.

Parameters	Training dataset
T_1 (°C)	$150 + 20 U(0,1)$
T_2 (°C)	$170 + 20 U(0,1)$
Rate (°C /min)	$1 + 3 U(0,1)$
1 st dwell duration (s)	$7200 U(0,1)$
Initial degree of cure (%)	$0.013 + 0.035 U(0,1)$

The two validation sets contain 1200 points each and include regular ranges of input parameters and extreme input variables within the range of the training set. These are summarised in Table 5. The regular set can provide information about how well the trained ML model performs on data which are generated for randomly different input parameter values.

Table 5: Validation datasets with different input parameter ranges; $U(0,1)$ denotes a uniformly distributed random variable in $[0,1]$.

Parameters	Regular set	Extreme set
T_1 (°C)	$150 + 17.5 U(0,1)$	150 points: $167.5 + 2.5 U(0,1)$
T_2 (°C)	$170 + 20 U(0,1)$	150 points: $167.5 + 2.5 U(0,1)$ 150 points: $190 + 5 U(0,1)$
Rate (°C /min)	$1 + 3 U(0,1)$	150 points: $0.75 + 0.5 U(0,1)$ 150 points: $4 + 0.5 U(0,1)$
1st dwell duration (s)	$7200 U(0,1)$	150 points: $7200 + 600 U(0,1)$
Initial degree of cure (%)	$0.012 + 0.035 U(0,1)$	150 points: $0.008 + 0.005 U(0,1)$ 150 points: $0.048 + 0.005 U(0,1)$

The role of this set is to provide insights about the ML model ability to generalise to different combinations of input values, within the original ranges used in the training dataset. The second validation set involves one variable taking extreme values that could be marginally outside the original range, and the rest kept in the same range as in the regular set. This validation set aims to assess the ML model performance when the input values deviate from the regular ranges providing information about how robust the model can be in handling outliers or unusual values in a predictive manner. A fixed random seed was applied during this

sampling to ensure that the same parameter combinations, and thus the same discrete validation sets, can be reproduced exactly in future runs.

2.3. Machine learning implementation

Predicting continuous variables, such as cure time and temperature overshoot, is considered a supervised regression task. Typical methods for supervised learning include conventional regression, decision trees, support vector machines and regularisation [32]. The selection of a regression ML model is crucial for data analysis. The traditional linear model often fails in cases involving nonlinear effects. Decision tree methods automatise detection of nonlinear main effects through recursive partitioning. Another similar method that minimises the risk of overfitting is random forest, which builds numerous trees in bootstrapped samples and aggregates them through averaging, providing robust predictions. Overfitting arises when a model corresponds too closely to a particular dataset and fails to accurately predict events in new samples. Bayesian regression trees utilise an underlying probability model for the structure and likelihood of data in terminal nodes. Additionally, algorithms like XGBoost and Light Gradient Boosting are commonly used in this category, providing further options for predictive modelling. On the other hand, support vector machines aim to find the hyperplane with the maximum margin between the two closest points in space. In contrast, regularisation techniques, like Elastic Net and LASSO, penalise models for increasing complexity, reducing the risk of overfitting by discouraging a large number of predictors compared to observations or a large number of collinear predictors.

To select the most appropriate ML algorithm for predicting the relevant continuous variables, these algorithms have been modelled using Python libraries and tools. More specifically, Python version 3.9 and the NumPy (version 1.23), pandas (version 1.5), and scikit-learn (version 1.2) libraries were used as the primary tools for preprocessing and training the proposed models. The evaluation process was conducted to assess the ML model performance and determine the candidate model. Each algorithm (i.e., stochastic gradient descent (SGD), decision trees, random forest, XGBoost, LightGBM, Elastic Net, and LASSO) using different set of parameters – hyperparameters – was trained and evaluated using the same data and performance metrics. Table 6 summarises these ML models and their critical hyperparameters. It should be noted that effectiveness and performance can still be affected by a variety of factors. While some of them are prone to overfitting and have their own limitations, data quality, data size, hyperparameter tuning, model

interpretability, and problem complexity can all influence their capabilities. Also, to ensure reproducibility of all ML training, evaluation and hyperparameter tuning stages, fixed random seeds were applied for dataset splitting, cross-validation folds, and parameter grid exploration, guaranteeing identical results upon rerun on the same software and hardware environment.

Table 6: Overview of candidate ML algorithms and their hyperparameters [32].

Algorithm	Description	Critical hyperparameters
Decision Tree	Decision tree methods automate interaction and non-linearity detection	Maximum depth: maximum depth of the tree.
Random Forest	Random forests build numerous trees in bootstrapped samples and generate an aggregate.	Number of trees: number of trees in the forest. Maximum depth: maximum depth of each tree in the forest.
XGboost	Gradient boosting machine known for its strong predictive performance.	Learning rate: step size shrinkage to prevent overfitting. Number of estimators: number of boosting rounds (trees). Maximum depth: maximum depth of each tree.
LightGBM	Gradient boosting framework designed for efficiency and scalability.	Learning rate: step size used to prevent overfitting. Number of estimators: number of boosting rounds (trees). Maximum depth: maximum depth of each tree.
Lasso (Linear Regression with L1 Regularisation)	Linear regression with L1 regularisation, leading to sparse coefficients.	Alpha: Constant that multiplies the L1 term, controlling regularisation strength.
ElasticNet (Linear Regression with L1 and L2 Regularisation)	Linear regression with a combination of L1 and L2 regularisation.	Alpha (Mixing parameter penalty): combination of L1 and L2 regularisation penalties. Lambda (Regularisation strength): overall strength of the regularisation.
Stochastic Gradient Descent (SGD) Regression	Linear regression with a stochastic gradient descent optimisation algorithm.	Learning rate: step size shrinkage used to prevent overfitting. Regularisation parameter: strength of the regularisation. Number of epochs: number of iterations through the training data.

To implement a working ML solution, it is necessary to explore its variables and parameters. This is achieved by following the steps outlined in Figure 4(a). The original dataset obtained from cure

simulation was inspected to gain insights into its characteristics, investigate the presence of outliers and evaluate the overall quality of data (Phase I). Additionally, apart from data quality assessment and ensuring consistency, the feature space within the dataset was explored to identify relevant features that could potentially influence the target variable or contribute to predictive modelling (Phase II). Subsequently, the simulation input values were used as input ML variables based on the previous analysis and were then fed into multiple ML models for the training and performance testing (Phase III). The XGBoost and LightGBM were chosen as the most suitable models by considering not only the predicted results but also the necessity to capture non-linear relationships, and the optimisation phase to fine-tune their performance was carried out (Phase IV). Following this phase, the model performance was tested against various validation datasets to ensure the robustness of the ML model and a better understanding of its capabilities and limitations (Phase V).

Figure 4(b) shows the integration of datasets in the overall procedure followed. Specifically, the FE model inputs and outputs were utilised to acquire any initial trends or patterns observed in the simulation process. Establishing good data quality is critical for the formulation of accurate and robust ML models for prediction tasks. The first step in this process was to assess the data quality of the ML model and evaluate its use for ML applications using EDA (Exploratory Data Analysis) (Phases I-II). Data integrity checks are performed to detect missing values, inconsistencies, or numerical anomalies. Outlier detection is typically conducted using both statistical thresholds (e.g., values exceeding ± 3 standard deviations from the mean) and interquartile range (IQR) methods. Physically unrealistic values—those violating known thermochemical constraints from the cure kinetics—were removed. This ensured that retained data reflected physically valid simulation behaviour. An important aspect of developing a reliable prediction model is understanding the importance of each feature or variable to the model, including model simplification, exploration, and the identification of the most important variables. To accomplish this, an initial analysis of feature importance on a cure dataset was performed. The main dataset attributes for the ML pipeline were the two dwell temperatures T_1 and T_2 , rate, 1st dwell duration and initial degree of cure targeting the prediction of cure time and temperature overshoot.

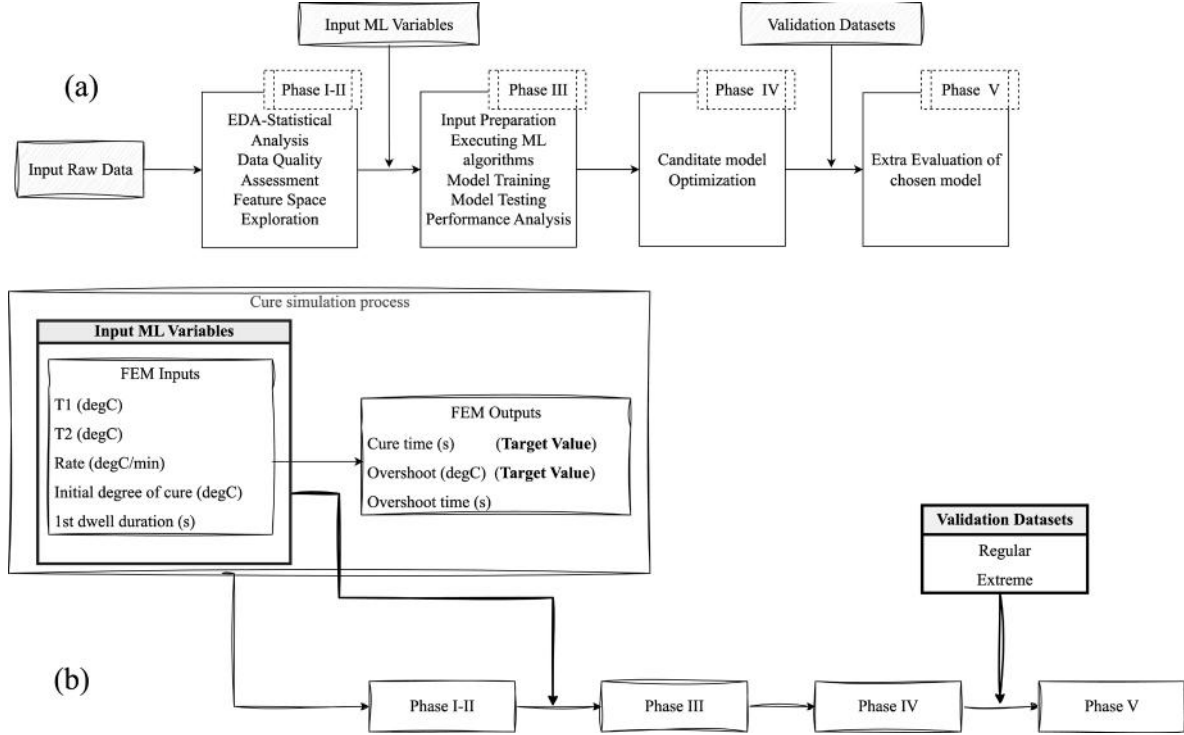


Figure 4: (a) Phases followed to build a robust ML model. (b) ML process: detail on input values/datasets.

As a part of Phase II, correlation techniques were employed utilising a tabular representation of correlation coefficients between variables. In this table, each cell represents the correlation between two variables. The correlation matrix was used to summarise data and provide valuable insights into relationships between simulation variables at this early stage.

To prepare the data for the training stage, it was necessary to perform feature scaling, which involved standardising the features by subtracting their means and scaling to unit variance (Phase III). This step is crucial for improving the accuracy and the convergence speed of the model. The training dataset was split into two sets, with 85 % used for training the models and the remaining 15 % forming the test set during training. The fixed test set was kept separate from the training process and used to acquire initial performance metrics, such as accuracy and error, for a diverse set of ML models. During the training stage, the initial parameters (hyperparameters) of each model were obtained and then tested on the fixed test set to acquire some initial performance metrics for minimising the error. The continuous target values represent a supervised regression task with the goal of minimising the loss function in the training process:

$$L = ||y - \varphi(x, \theta)|| \quad (13)$$

Here L is the loss function calculated as the absolute difference between the true value y and the predicted value $\varphi(x, \theta)$. In training, the objective is to reduce this loss function by modifying the model parameters represented by vector θ . The squared loss function is the squared difference of Eq. (13) and is known as Squared Error Loss. The cost function for the Absolute Error or L loss, is known as the Mean Absolute Error (MAE), where the cost function for Squared Error is called Mean of Squared Error (MSE).

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (14)$$

Here N is the number of data samples, y_i is the true value and $\hat{y}_i$ is the predicted value. Hence, the square root of MSE is used to calculate the root mean squared error (RMSE) which accounts for variations from the actual value and measures the average magnitude of errors.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (15)$$

Aside from the metrics listed above, 95% confidence intervals (CIs) were calculated for the square root of MSE, RMSE and MAE for the target variables. The CIs were obtained using non-parametric bootstrapping with 1000 resamples of the residuals of each dataset. For each bootstrap sample b , the performance metric $\hat{\theta}_b^*$ was recalculated, and the CI bounds are defined as:

$$CI_{95\%} = [\hat{\theta}_{(2.5\%)}^*, \hat{\theta}_{(97.5\%)}^*] \quad (16)$$

where $\hat{\theta}_{(p\%)}^*$ represents the p th percentile of the bootstrap distribution [33].

Ensemble ML algorithms combine many base models to improve the performance and minimise the prediction errors of regression problems and feature selection. Specifically, Voting Regressors (VRs) utilise a weighted averaging of estimators from all models to produce a final estimate. An ensemble voting regression model was developed to estimate the cure time and temperature overshoot by employing LightGBM and XGboost regressors.

Both LightGBM and XGBoost are gradient boosting algorithms that utilise an ensemble of regression trees to achieve fast learning speed and accurate prediction of continuous target values. LightGBM is a widely used boosting algorithm known for its efficiency and flexibility [34]. For supervised learning problems, a system consisting of a random output variable y and a set of random input variables $\mathbf{x} = \{x_1, \dots, x_n\}$ using a training sample of $\{y_i, \mathbf{x}_i\}_1^N$ of known $(y, \mathbf{x})$ - values targets to obtain an estimate $\hat{f}(\mathbf{x})$ of the function $f^*(\mathbf{x})$ mapping $\mathbf{x}$ to y . The goal of the LightGBM algorithm is to minimise the expected

value of a particular loss function L by finding an approximation of $\hat{f}(x)$ of a function $f^*(x)$ over the joint distribution of all $(y, \mathbf{x})$ - values.

$$\hat{f} \sim f^*(x) = \operatorname{argmin}_f E_{y,x} L(y, \varphi(x, \theta)) \quad (17)$$

Specifically, the input values of the FE model were fed into 40 ML models (as shown in Table 7) setting their hyperparameters defining the one-by-one structure of each model. Table 7 presents 28 of the 40 ML models that were evaluated and assessed based on their performance metrics for the prediction of cure time. To ensure robustness and optimal hyperparameter tuning of an ML model, a k-fold cross-validation approach combined with a grid search technique was used. The k-fold method allows to validate the model performance across multiple subsets of the data, while the grid search approach enables to systematically explore a range of hyperparameters and identify the optimal combination for maximising predictive accuracy. To ensure a fair comparison and assess the efficiency of these selected models, the 5-fold cross-validation technique was employed. This involves dividing the training dataset into 5 subsets, and each model undergoes five iterations of training and validation, with different subsets serving as the validation-test set in each iteration. The 5-fold cross-validation was repeated k times ($k = 5$), providing a stable and reliable evaluation of the model's generalisation capabilities.

In Phase IV, the LightGBM and XGBoost models were chosen based on their higher prediction accuracy and lower RMSE compared to the other ML algorithms and set to their final optimisation using the grid-search technique as an exhaustive search algorithm for finding optimal hyperparameters. An optimisation method was used based on the testing dataset to determine the best hyperparameters for the models to achieve the best accuracy. The efficiency of these two proposed models was validated also through 10-fold cross-validation. This technique divided the dataset into 10 subsets, and each model was trained and validated ten times, with a different subset as the validation set in each iteration. The 10-fold cross-validation ensured robust evaluation, reducing overfitting and bias [35]. The selection of XGBoost and LightGBM was also based on the characteristics of the dataset and the modelling objectives. The cure process data are structured, tabular, and moderately sized ($\sim 10^4$ points) with multiple interacting process parameters. Both XGBoost and LightGBM offer scalability and robust handling of multicollinearity and missing values and provide feature importance metrics that facilitate interpretability—an important consideration for industrial adoption. Their computational efficiency enables inference, making them more practical for real-time optimisation than deep learning frameworks, which often require orders of magnitude

more data and training time for comparable accuracy. These algorithms have also demonstrated strong performance in recent manufacturing and composites process modelling studies [36,37].

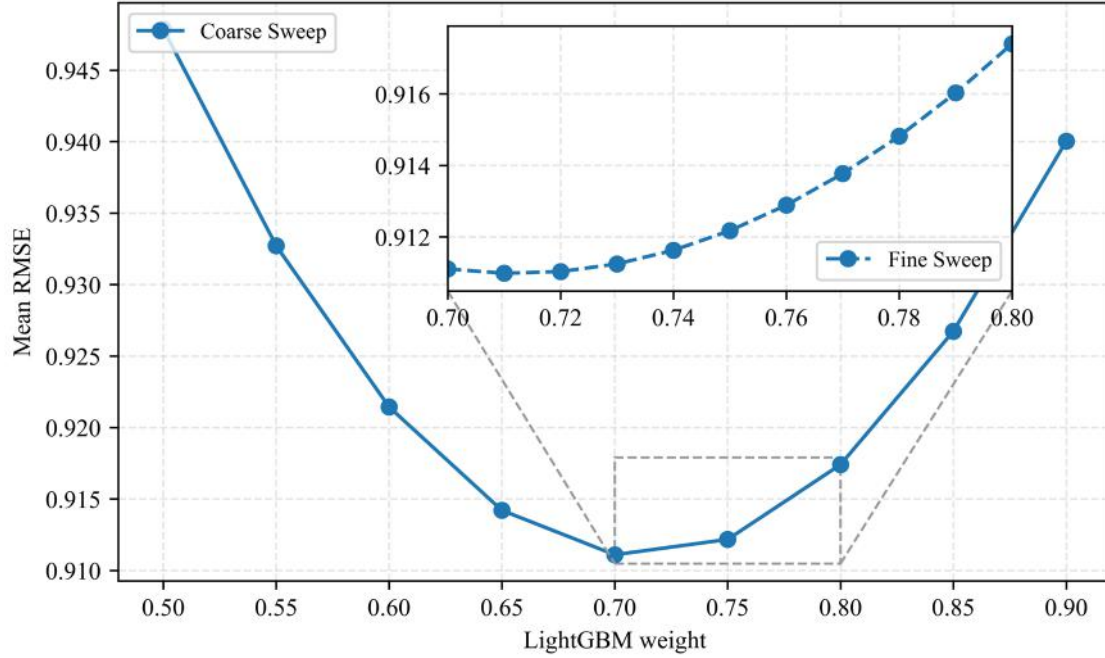


Figure 5: RMSE trends for coarse (solid) and fine (inset, dashed) weight sweeps of the LightGBM XGBoost Voting Regressor.

Following this individual optimisation, LightGBM and XGBoost were integrated into a Voting Regressor to leverage their complementary strengths. Ensemble weights were tuned via a two-stage grid search with 5-fold cross-validation. In Figure 5, a coarse sweep (0.50–0.90, step 0.05) revealed a broad plateau between 0.70–0.75 LightGBM weight (mean $R^2 \approx 0.998$; RMSE ≈ 0.911). A fine sweep (0.70–0.80, step 0.01) confirmed that the selected 0.76/0.23 configuration lies within this plateau, with RMSE differences less than 0.002 from the best combination—well within cross-validation variability. Such flat optima are common when base learners exhibit correlated error patterns [38,39]. The small deviation from a total weight of 1.0 is due to floating-point precision.

Table 7: Overview of 28 of 40 regression models indicating some of the different hyperparameters.

Algorithm	Normalised RMSE	Hyperparameters
VotingEnsemble	0.00754	algorithm: ['LightGBM 'XGBoostRegressor']
MaxAbsScale, LightCBM	0.00830	min_data_in leaf: 20
MaxAbsScaler, XGBoostRegressor	0.00999	tree_method: auto
MaxAbsScaler, ExtremeRandomTrees	0.02442	bootstrap, criterion: mse,max_features =0.8
RobustScaler, LightGBM	0.02977	bootstrap, criterion: mse,max_features =0.7
MinMaxScaler, RandomForest	0.03152	bootstrap, colsample=1, learning_rate=0.01
StandardScalerWrapper, ExtremeRandomTrees	0.03386	bootstrap, criterion: mse,max_features =sqrt
StandardScaledWrapper, RandomForest	0.03471	bootstrap, criterion: mse,max_features =0.7
Min MaxScaler, ExtremeRandomTrees	0.03571	bootstrap, criterion: mse,max_features =0.4
MinMaxScaler, LightGBM	0.03954	boosting_type=gbdt, colsample_bytree=0.8
RobustScaler, RandomForest	0.04030	bootstrap, criterion : mse, colsample_bytree=0.8
RobustScaler, ExtremeRandomTrees	0.04544	bootstrap,criterion : mse,max_features =0.9
Min MaxScaler, LightGBM	0.04823	bootstrap,criterion : mse,max_features=0.7
MinMaxScaler, ExtremeRandomTrees	0.04914	bootstrap,criterion: mse,max_features=0.5
RobustScaler, DecisionTree	0.05472	criterion: friednan_mse, min_sample_leaf=0.00236
MinMaxScaler, DecisionTree	0.05649	criterion: friedman_mse, min_sample_leaf=0.0034
StandardScalerWrapper, ExtremeRandomTrees	0.05652	bootstrap,criterion : mse,max_features =0.7
Min MaxScaler, LightGBM	0.06434	boosting_type: gbdt,colsample_bytree=0.6
StandardScalerWrapper, RandomForest	0.06727	bootstrap, criterion: mse,max_features = 0.4
MaxAbsScaler, ElasticNet	0.06767	alpha=0.001, l1_ratio=1
StandardScalerWrapper, ElasticNet	0.06767	alpha=0.001, l1_ratio = 0.84368
MinMaxScaler, ExtremeRandomTrees	0.06894	bootstrap,criterion: mse,max_features =0.4
RobustScaler, ElasticNet	0.06964	Alpha=0.001,l1_ratio : 0.68736
TruncatedSVDWrapper, LassoLars	0.08850	alpha=0.47421,normalise
MinMaxScaler, SGD	0.09781	eta0=001, fit_intercept, l1_ratio=0.2244
StandardScaledWrapper, SGD	0.12553	eta0=001, fit_intercept, l1_ratio=0.3469
Min MaxScaler, SGD	0.21921	epsilon=0.18776,eta0=0.0001
MinMaxScaler, SGD	0.21967	epsilon=0.004179,eta0=0.0001

Moving to Phase V, the selected LightGBM and XGBoost models were evaluated by incorporating the validation set to ensure their robustness and reliability. The extra evaluation of the candidate model involved the utilisation of two validation datasets, which are detailed in Table 3. The validation sets -regular

and extreme - which were distinct from the training sets, served as an independent benchmark for evaluating the performance of the ML models. Unlike the iterative evaluation in previous phases, where 5-fold and 10-fold cross-validation were used to assess the model's generalisation capabilities, the validation sets in Phase V were introduced to rigorously validate the models on entirely new and unseen data that simulate challenging scenarios. This evaluation process helps to gain an in-depth understanding of predictive power since it goes beyond the training feature space and validates the model ability to generalise to diverse feature spaces and potential outliers.

2.4. Process trials

The study initially utilised a 12 mm thick composite plate for training and research purposes. To further confirm the practical relevance and robustness of our approach and verify the resilience of the methodology and replicate actual manufacturing conditions, two experimental trials were conducted using a 3 mm thick composite part. The main goal was to establish a complete understanding of curing processes and predictive modelling under complex and challenging conditions common in industrial applications. The ML models were retrained using FE data from the 3 mm composite part, following the same process as the one described previously for the 12 mm plate. The low thickness of this component makes the problem simpler. Therefore, the Results and Discussion section focuses on the richer trade-off observed in the 12 mm case.

The input variables presented in Section 2.2 were utilised to define experimental heating cycles for the process trial carried out using an ISOJet single part RTM system connected to a self-heated tool (Figure 6). The composite tool was used to manufacture flat test laminates 400 mm × 300 mm and 3 mm thick, consisting of a $[0,90]_s$ layup of the HTA40 carbon reinforcement (Hexcel G0926 fabric). The composite tool comprises a top plate (punch) and a bottom plate (matrix). Both plates are equipped with internal heating elements, and each is divided into six heating zones, as shown in Figure 6(b). Two k-type thermocouples are embedded in each zone to monitor and control the curing cycle. The average temperature of the two thermocouples is used as the control signal and each of the twelve zones independently, with a common set point used in all zones to follow the cure profile. The material is placed in the area of zones 2 and 3, with the rest of the zones controlled to the same temperature profile to minimise in-plane heat losses and ensure one-dimensionality of the heat transfer phenomena.

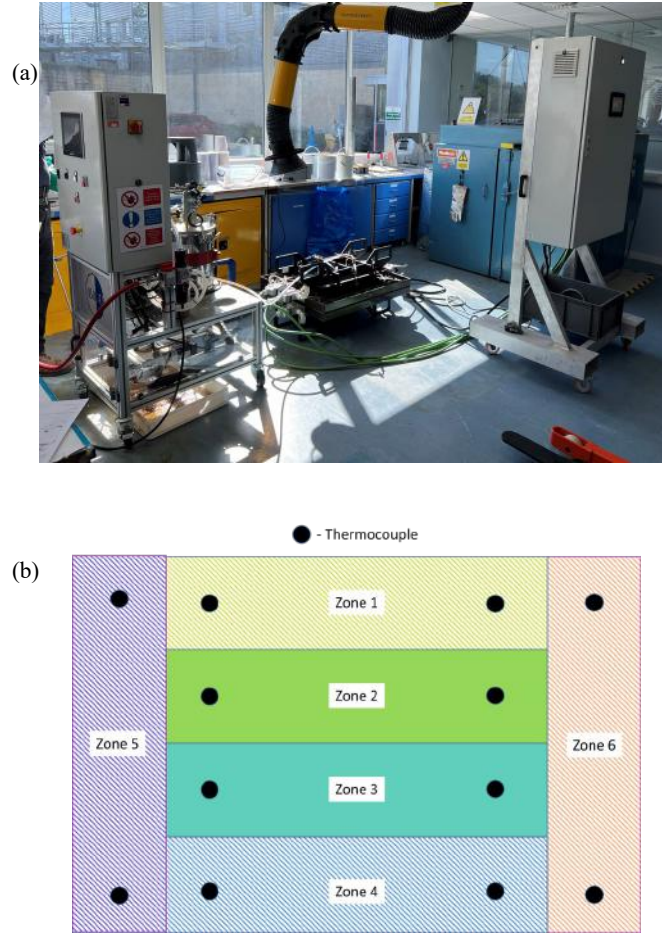


Figure 6: (a) RTM process equipment setup using ISOJet RTM Injection Machine (left), composite tool (centre, floor) and control box (right) for tooling heating, monitoring and control; (b) in plane schematic of heating zones on one side of the tool.

The RTM process parameters prior to the cure cycle were kept constant to remove the influence of process parameters outside those being studied. The RTM6 resin was allowed to fully thaw to room temperature over 48 h prior to being degassed at 80 °C and 0.5 bar vacuum for 45 min. The resin was then transferred to the ISOJet system and placed in the reservoir and preheated to 80°C. The injection and return hoses were connected to the composite tool which was placed under a vacuum of 0.32 mbar and preheated to 120 °C. The inlet hose was preheated to 110 °C, and the resin injected at 5 bar. During injection, the maximum flow rate was set at 150 cm³/min. Once fully injected, with resin clearly showing at all outlets, these were closed off and the injection pressure increased to 6 bar. At this point the selected cure cycle was started and the panel cured. On completion of the cure, the laminates were demoulded from the tool. A composite plate was produced based on two dwell profiles, where the predicted cure time was used instead of the simulation cure time to consider the part cured and terminate the curing process. Table 8 lists the specific curing profiles for the composite part.

DSC was employed to assess the outcome of process trials. A DSC 250 Discovery (TA Instruments) was used equipped with an RCS cooling system under nitrogen purge at 50 mL/min. For each experiment, grounded samples of the cured composite with a weight of 6-7 mg were encapsulated in a Tzero aluminium pan and scanned at 10 °C/min from 70 °C to 230 °C to measure the uncured material glass transition, exothermic cure onset temperature, cure peak temperature and heat of reaction. The glass transition temperature was identified using the inflection point of heat flow versus temperature curves.

Table 8: Experimental trial using heating cycle based on predicted cure time.

	T ₁ (°C)	T ₂ (°C)	Ramp (°C/min)	1 st dwell duration (min)	2 nd dwell duration (min)
Trial 1	151	170	1.2	75	17
Trial 2	152	170	2.4	5	57

3. Results and discussion

3.1. Datasets

Figure 7(a) illustrates the exhaustive search range with respect to the total cure duration and overall temperature overshoot objectives. The cure duration ranges from about 2500 s to 10000 s, whilst the overshoot from 2.5 °C to 12 °C. There is a Pareto front of solutions which can be divided into two regimes: one of conservative profiles which combine a relatively long duration of 4000–7000 s with a low overshoot in the range of 2.5-3 °C and one of faster profiles that combine relatively short durations in the 2500-4000 s range with greater overshoots of 3-12 °C. These two regimes are representative of the trade-off addressed in most curing problems of non-thin parts. Process efficiency is accompanied by increased risk of exothermic effects and safe profiles lead to long and expensive long process durations. However, within the range of each regime there are substantial optimisation opportunities, e.g. there are many cure profiles in the corner region of the L-shaped Pareto front which result in cure durations in the 3500-4500 s range combined with relatively low overshoots of 2.5-4 °C.

Figure 7(b) illustrates a comparison between the overall training set and the two independent validation tests. It can be observed that the validation sets cover the whole range associated with the two primary process metrics. In the case of the set including extreme values of input, there are few points that are slightly outside the envelope of solutions provided in the training set as a result of the expansion of the

input range of this set, which aims to challenge the ML method. Overall, for all validation sets, there is greater density of points in the lower overshoot/greater duration part of the space. This is a reflection of the density of points in the overall dataset, produced by the way the input variables are mapped onto the curation-temperature overshoot space.

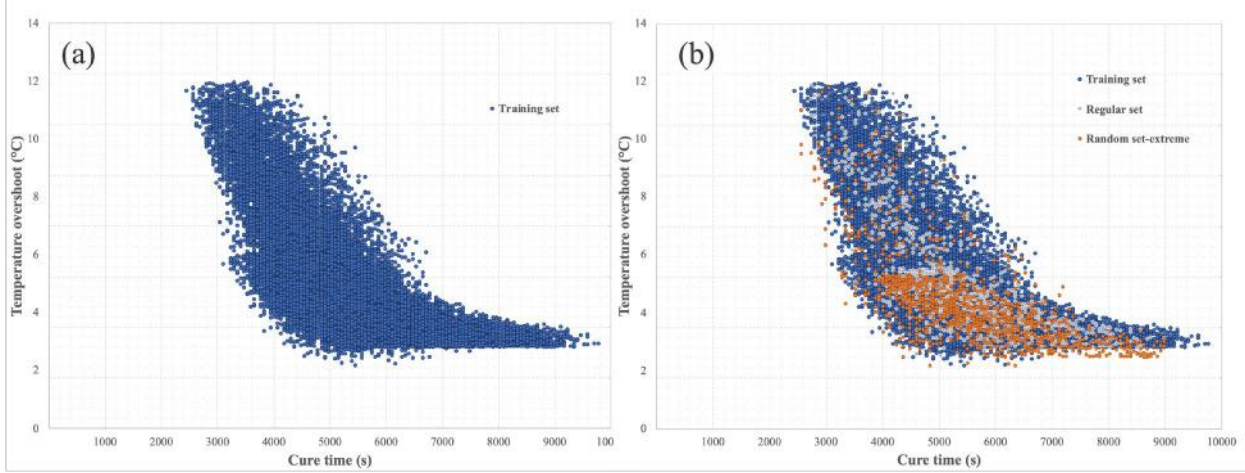


Figure 7:(a) Temperature overshoot vs cure time for training set, (b) Temperature overshoot vs cure time for regular and extreme set overlaid on training dataset.

3.2. Statistical Analysis

Table 9 summarises the findings of the data quality analysis. The high data quality indicates the suitability of the set for ML as there are few missing or erroneous values (less than 1 %) and a statistically normal distribution. In simulation analysis, the narrow range and relatively small standard deviation values for both T_1 and T_2 , as well as the large spread and notable standard deviation values for the first dwell duration are a reflection of the choices made for the dataset. The range selection of input values is a decision that can be adjusted based on the specific requirements and objectives of the study. The duration of cure and overshoot play a crucial role in the curing model behaviour and performance. As outputs, they are indicators of how the curing process behaves under different conditions and understanding their variation offers valuable insights for process optimisation and control. The moderate standard deviation value for overshoot implies some variation in the occurrence of overshoot during cure.

Table 9: Descriptive statistics, part of EDA (Phase I-II), on the 1-D curing model.

Parameters	N	mean	std deviation	min	max
T ₁ (°C)	32768	160.00	5.79	150.00	170.00
T ₂ (°C)	32768	180.02	5.77	170.00	190.00
Rate (°C/min)	32768	2.51	0.87	1.00	4.00
1st dwell duration (s)	32768	3578.3	2075.9	0.0	7200.0
Initial degree of cure	32768	0.0306	0.0101	0.0130	0.0480
Cure time (s)	32768	5259.8	1162.4	2460.0	9780.0
Overshoot (°C)	32768	5.17	1.87	2.15	11.92

Based on the data obtained from the correlation table (Table 10) and in conjunction with inputs from cure simulation, an investigation was conducted to explore the important variable relationships for the model and gain physical insights. The correlation table represents the potential relationships between variables as well as the strength and direction of associations between the features. The correlations between input variables are negligible showing that the sampling used was successful in terms of independence of variables. This low multicollinearity benefits ML by reducing feature redundancy and improving model stability. There is a strong correlation of 0.69 between 1st dwell duration and cure time. This is a causal effect as increasing the dwell duration directly increases total process time. Cure time is negatively correlated (-0.47) to the 1st dwell temperature as the Arrhenius dependence of reaction rate on temperature means that curing progresses further as the temperature increases leading to shorter process duration. The temperature overshoot is negatively correlated to the 1st dwell duration (-0.62). A shorter 1st dwell causes greater remaining reactivity as temperature increases towards the 2nd dwell resulting in greater overshoot. Similarly, the weak negative correlation of overshoot to 1st dwell temperature (-0.26) is due to the fact that a lower temperature causes greater remaining reactivity. The correlation between cure time and overshoot (-0.74) is non-causal, as milder conditions resulting in longer curing times also result in smaller overshoot magnitudes. These insights provide valuable information for understanding the curing process and optimising reaction conditions, since the correlation matrix analysis contributes to a deeper understanding of the complex interactions between variables in the curing process.

Table 10: Correlation table.

	T₁	T₂	Rate	Δt_1	Initial cure	Cure time	Overshoot
T₁	1	0.0025	0.0011	0.00037	0.0054	-0.47	-0.26
T₂		1	-0.0085	0.001	0.0015	-0.099	-0.12
Rate			1	-0.01	-0.009	-0.3	0.17
Δt_1				1	0.0016	0.69	-0.62
Initial cure					1	-0.2	-0.15
Cure time						1	-0.74
Overshoot							1

3.3. Overshoot prediction

Targeting the overshoot values, the proposed Voting Ensemble model used the LightGBM and XGBoost models as ensemble members with 0.76 and 0.23 respectively for the weighted average overshoot predictions. The model explains a very large proportion of variation in the overshoot testing dataset ($R^2=0.983$) and model consistency is characterised by a low MSE at 0.058 and MAE at 0.099. As depicted in Figure 8(a), there is a linear correlation between prediction and FE results which is very close to the diagonal line (45°), which confirms that the correlation coefficient is very strong for the testing set (15 % of the training set).

Table 11: Overshoot prediction results: Detailed presentation of performance metrics on training and testing datasets, and validation with regular and extreme sets.

Performance metrics	Training set	Testing set	Regular set	Extreme set
R^2	0.99	0.9833 (0.979 – 0.986)	0.986 (0.983 – 0.989)	0.958 (0.950 – 0.965)
RMSE	0.18	0.24 (0.23 – 0.26)	0.23 (0.22 – 0.25)	0.41 (0.39 – 0.44)
MSE	0.03	0.058	0.05	0.17
MAE	0.09	0.099 (0.094 – 0.104)	0.11 (0.10 – 0.12)	0.18 (0.17 – 0.19)

Figure 8 presents the correlations with slopes of 0.99 and 0.98 for FE vs predicted overshoot values of regular and extreme datasets. The predicted overshoot values for the regular model are shown in Figure 8(b) with R^2 of 0.986 for the regular and 0.958 for the extreme validation set (Figure 8(c)). The extreme dataset includes the greatest deviation from the actual values as expected since highest values of overshoot

were present. Considering the performance metrics of each dataset (Table 11) the overall predictions exhibit a similarity to the target values achieving up to 95 % in predicting overshoot. The bootstrap 95 % confidence intervals for RMSE and MAE are small relative to their means, confirming low variability in prediction error. Whilst performance decreases slightly for the extreme set, the CIs remain narrow, suggesting that this drop reflects the inherent difficulty of extrapolation rather than instability in the model.

A better visualisation of the pattern of differences between the observed values and the predicted values can be shown by the distribution of residuals illustrated in Figure 8(d). The residuals for each validation dataset are normally distributed around zero; It is worth noting that the overshoot values of the Voting Ensemble model match the simulation output values very closely. The fact that the residuals are centred around zero further supports the notion that the model is unbiased and provides reliable predictions. An approximate range of $[-0.2, 0.2]$ for all validation model residuals suggests that the smaller the range of residuals, the better the model predictive performance as it indicates that the model is making precise predictions with minimal deviation from the true values.

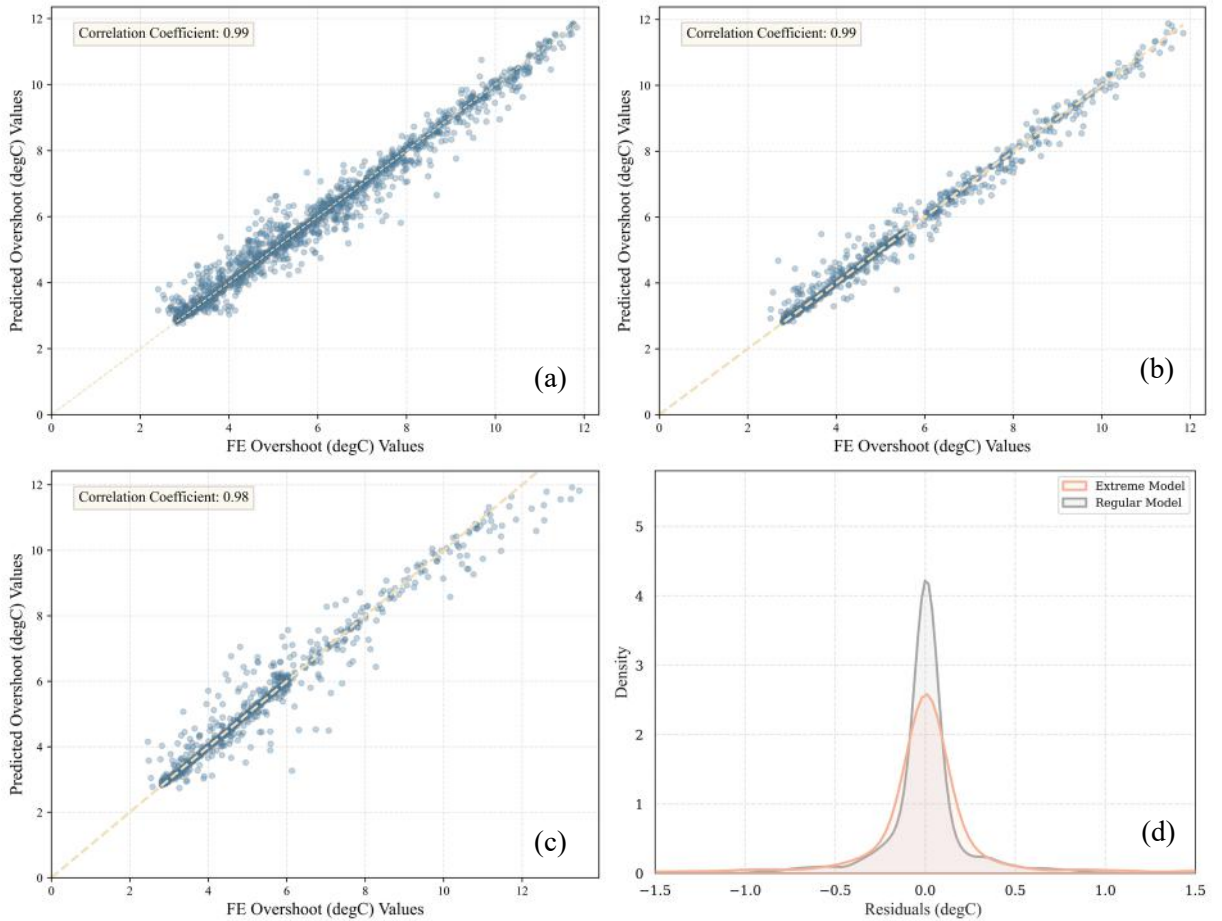


Figure 8: Scatter plot of predicted and FE overshoot values on the (a) test dataset, (b) regular and (c) extreme dataset, (d) Distribution of overshoot residuals for two validation datasets.

3.4. Cure time prediction

The optimisation of the Voting Ensemble model yielded the most precise predictions when the ensemble weights were 0.75 and 0.25 for the LightGBM and XGboost regressor. Figure 9(a) shows the scatter plot of the testing set for cure time where R^2 is 0.999 and the performance measurements (MSE and MAE) validate the model accuracy with extremely low values. From Figure 9(b) and (c), it is apparent that the Voting Ensemble model demonstrates strong convergence since the correlation coefficient is almost 1 for all datasets. More specifically, the scatter plot of regular model reveals a very clear linear trend with tight clustering of data points. The R^2 of the regular set shows that the ML predictor can explain 99 % of the variation in the cure time target variable. Considering that the most extreme values of overshoot affect the cure time process, the R^2 score of 0.984 indicates a strong correlation between the predicted and FE values. Additionally, as shown in Table 12, the RMSE and MAE values are relatively low, suggesting that the model is able to make accurate predictions with minimal error. This demonstrates the effectiveness and robustness of the model in predicting cure time values under various conditions. The 95% CIs for R^2 , RMSE, and MAE are narrow, indicating stable performance and minimal sensitivity to sampling variation.

As observed in Figure 9(d), the residuals of the regular model are approximately between -0.14 and 0.14 which represents a small specific interval within which the majority of the residual values fall. The offset of 0.05 in the distribution of residuals of the extreme model is desirable as it indicates that the model is making accurate and precise predictions. The small offset implies that the model is capturing the underlying patterns in the data effectively and is not systematically overpredicting or underpredicting the target variable.

Table 12: Cure time prediction results: Detailed presentation of performance metrics on training and testing datasets, and on regular and extreme validation sets.

Performance metrics	Training set	Testing set	Regular set	Extreme set
R^2	0.99	0.99 (0.989 – 0.991)	0.984 (0.981 – 0.986)	0.998 (0.996 – 0.998)
RMSE	0.03	0.04 (0.038 – 0.042)	0.16 (0.15 – 0.17)	0.15 (0.14 – 0.16)
MSE	0.001	0.002	0.026	0.023
MAE	0.029	0.035 (0.034 – 0.036)	0.13 (0.12 – 0.14)	0.13 (0.12 – 0.14)

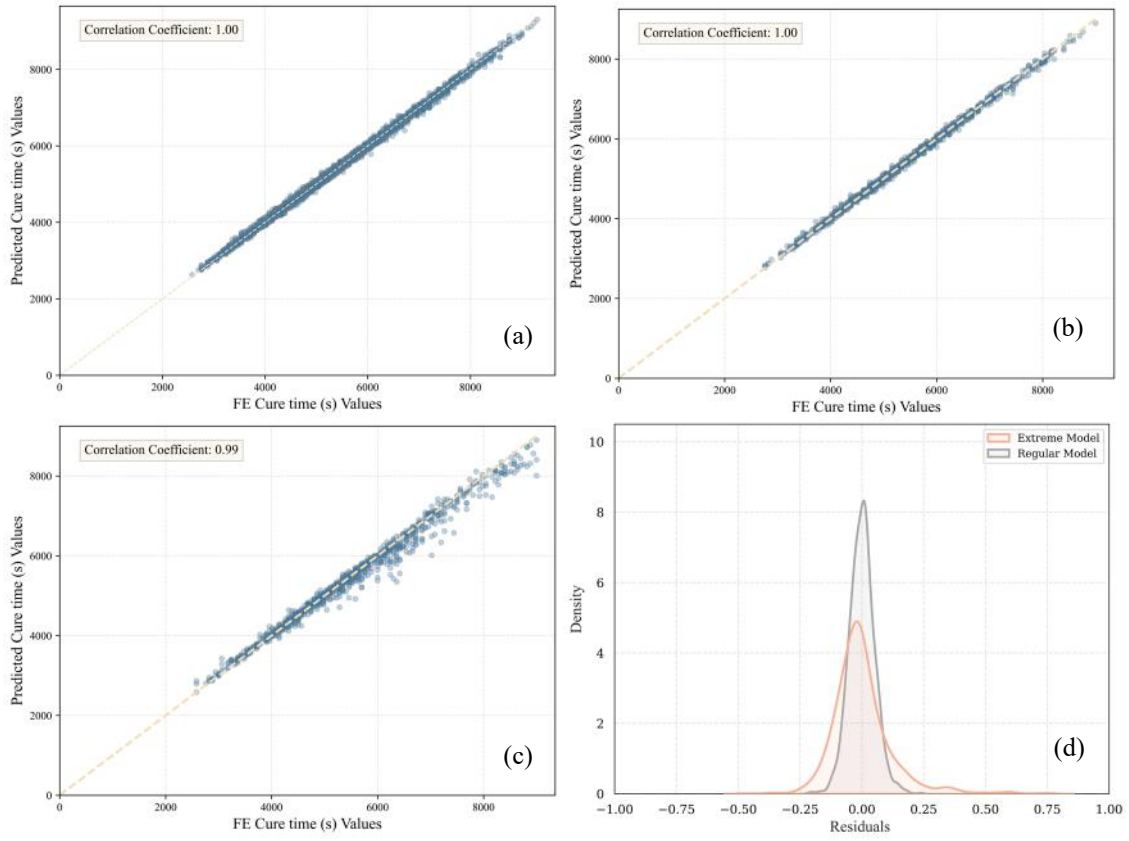


Figure 9: Scatter plot of predicted and FE cure time (s) values on the (a) test dataset, (b) regular and (c) extreme dataset. (d) Distribution of cure time residuals for two validation datasets.

3.5 Cross correlation prediction

The results reported in Section 3.4 demonstrate the high accuracy of ML models in independently predicting the variables temperature overshoot and cure time. However, the accuracy evaluation is valuable not only for testing the ML models predictive performance but also for understanding their ability to capture physical interactions. Optimising the cure is a trade-off problem. Making the process shorter by using a more aggressive thermal profile has a penalty in terms of temperature overshoot and conversely, targeting low temperature overshoots comes at a cost of longer process processing cycles. A metamodel, such as the ML solution proposed in this work, must reproduce the cross-correlation between the two critical variables - cure time and temperature overshoot - to be able to capture the trade-off between them when optimising the process design. Figure 10 shows whether the models account for cross-correlation between target values. Specifically, the graph of temperature overshoot against cure time for both predicted and actual values from the training set can visually confirm that the model reproduces the patterns of correlation between these two target outputs. The overall picture is confirmed by the correlation coefficients between

the two variables for the ML prediction and the FE model, which are 0.744 and 0.733 respectively, indicating the ML model successfully reproduces the correlation structure of the dataset. Furthermore, the ML solution set reproduces the Pareto front of the multi-objective problem of minimising the cure time and temperature overshoot. The small differences between FE and ML points in Figure 10, which are in the order of seconds, indicate that the predicted values of temperature overshoot and cure time are relatively close to each other for the actual corresponding data points. The Voting ensemble models maintain the consistency in recording the relationship between the predictive values. This level of consistency and correlation can form the basis of model-based control.

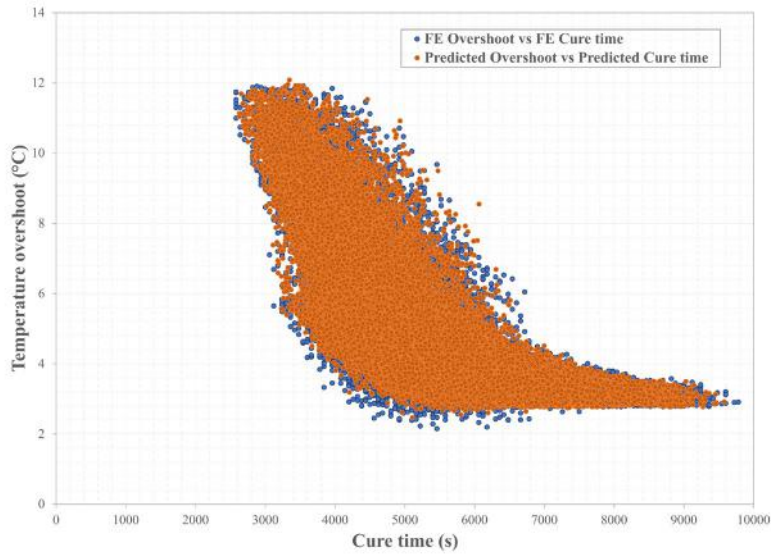


Figure 10: Relationship between overshoot and cure time: predicted vs. FE values.

3.6. Model execution time and model contribution

One of the challenges in the development of digitisation schemes for composite manufacturing is the high complexity and computational time of FE simulations [9,10]. Even running 1-D models can take 1-2 minutes to solve for a single data point which result in hours or days for an optimisation task. This increased time per point explains why the proposed models for the two target values can offer advantages over traditional FE simulation, in terms of computational cost while maintaining a high level of accuracy.

The algorithm was implemented and run using Python version 3.9 on a computer with M1 chip with 200 GB/s memory bandwidth and 16 GB of RAM. Table 13 compares the execution time of the models to FE simulation for each target value. The ML model can be executed for the two variables in less than 30

ms. This means that an optimisation task requiring a few thousands function evaluations (e.g. a GA with 50-100 individuals needing a few tens of generations to converge) can be completed with 2-3 minutes making it feasible to make it part of the control loop.

Table 13: Execution time for ~ 1000 samples.

Target value	Average ML execution time (s)	Average FE execution time(s)
Overshoot (degC)	0.0131	90 (calculated simultaneously)
Cure time (s)	0.0142	

Beyond GA approaches, the proposed ensemble method was benchmarked against alternative ML methods within this study (Table 7), including gradient boosting, bagging ensembles, kernel methods, and regularised regressors. Recent literature reports the application of deep learning frameworks—CNNs, physics-informed neural networks and conditional generative adversarial networks—for predictive modelling of cure-related phenomena, residual stresses, and thermal fields in composites [16,40]. While these methods show strong predictive potential, they often require training datasets of 10^5 – 10^6 samples, GPU-accelerated training lasting several hours to days, and inference times exceeding 0.5–1 s when coupled with physics solvers, making seamless real-time deployment challenging.

The proposed voting-ensemble method achieves competitive predictive accuracy using a relatively small dataset ($\sim 10^4$ FE simulations), trains in minutes on a standard CPU, and has a compact memory footprint suitable for embedding in process-control software. Previous studies have also shown that XGBoost can outperform other ensemble methods such as AdaBoost and GBM in manufacturing applications [15], due to its capacity to handle missing values, built-in regularisation, and ability to capture complex nonlinear process relationships.

3.7. Process trials results

Table 14 reports the results of the ML-based and FE cure time predictions from the 3 mm panel process trials. In these experimental trials, the ML model showed a strong alignment with the FE model. The difference in cure time results corresponds to a small error of 0.6-3.1 %, demonstrating the convergence of these two approaches. Figure 11 illustrates the DSC analysis for the trial laminates. The heat flow data reveal a near-optimal cure state with the glass transition point (T_g) at approximately 191-194 °C identified as the inflection point of the transition, following recommendations for composites [41] and using the

thermal analysis software TRIOS [42] to carry out the determination. This is in line with nominal values reported in the 194-210 °C range based on Dynamic Mechanical Analysis (DMA) measurements [[29], which tend to overestimate the value compared to DSC.

Table 14: Results of ML and FE predictions for the process trials.

	ML Predicted cure time (min)	FE cure time (min)	Relative Difference in cure time
Trial 1	132	128	3.1 %
Trial 2	82.5	82	0.6 %

Overall, the two physical process trials were chosen to balance the high cost and complexity of RTM processing with the need for representative validation. Given the low variability observed, the strong convergence between ML, FE, and DSC results, and the extensive prior validation of the ML framework using a large FE dataset, this sample size is sufficient for a first demonstration of the framework’s accuracy. Furthermore, given the narrow simulation confidence intervals and the observed trial deviations well within established process variation limits, the probability of a large undetected discrepancy can be considered low. These results serve as an experimental indication of the alignment between the ML produced cure cycles and experimental results. Furthermore, this enhances the practical value of the utilisation of the ML based approach, whilst the extrapolation of validity to greater thicknesses where exothermic effects play a significant role follows from the accuracy by which the ML metamodel represents the widely validated finite elements approach [4,43].

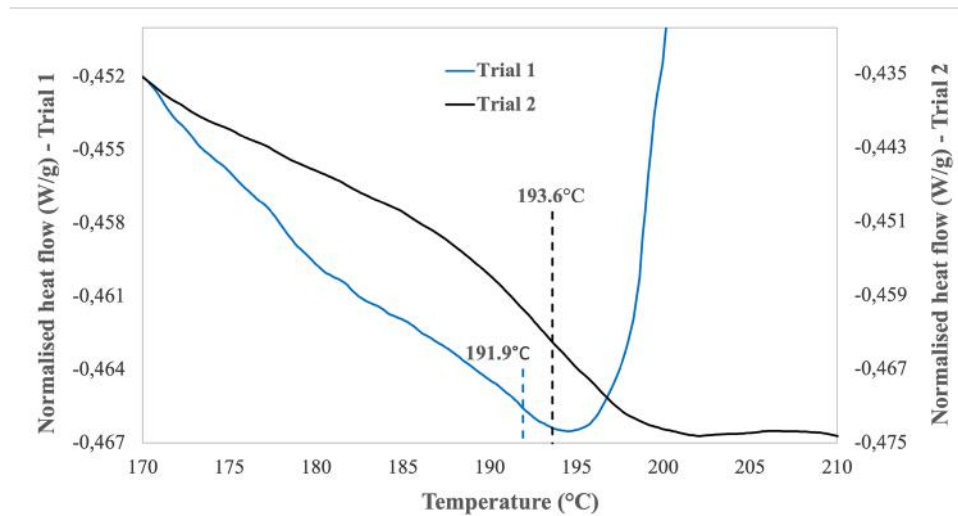


Figure 11: DSC results for the process trials and identification of glass transition temperature.

3.8. Industrial applicability and limitations of the modelling framework

The ML modelling approach developed here demonstrated the capability to simulate accurately the critical parameters of the thermochemical aspects of the cure process and their interactions. Beyond predictive accuracy, industrial deployment of the proposed framework requires trust from process engineers, particularly in safety- and quality-critical manufacturing contexts. Interpretability is essential to ensure a transparent mapping between input features (e.g., cure cycle settings, initial material state) and predicted outputs (e.g., cure time, temperature overshoot). Future developments will integrate explainable ML techniques—such as feature importance analysis, partial dependence plots, and SHapley Additive exPlanations (SHAP) values—to quantify the relative influence of each input feature on model predictions [44]. These methods, demonstrated in recent composites manufacturing studies, would enable engineers to verify model reasoning, identify key sensitivities, and adjust process parameters with confidence. Incorporating domain-specific knowledge through physics-informed learning will further align the model reasoning with established thermochemical curing dynamics, enhancing reliability in real-world manufacturing [18]. Addressing these interpretability aspects in parallel with predictive performance will support broader industrial acceptance and safe application of the model.

The methodology followed applies to thermosetting matrix composites in which one-dimensional heat transfer and through thickness thermal and material state gradients are the dominant phenomena during the process of cure. This corresponds to a great range of composite components in which in plane dimensions are significantly greater than the thickness and curvatures are moderate in comparison to the thickness (e.g. skins and plates). Components with three dimensional features (e.g. joints and areas with stiffeners) would require the incorporation of 3-D heat transfer effects in the training dataset through finite elements simulation of relevant geometries. Similar trade-offs between cure time and temperature overshoots exist in such geometries [23], which would be represented in an appropriate manner in the developed framework. However, the interplay between thermal gradients developed in the thickness direction and more effective heat dissipation in the fibre direction is likely to enrich the landscape of relevant outputs and their mapping onto process variables, requiring greater computational resources for training and potential modification of the ML metamodels. Similar effects might arise in tow dimensional geometries, e.g. composite cables and booms.

The material behaviour represented here corresponds well to the industrial state-of-the-art, covering fully epoxy based material as well as other thermosetting chemistries with kinetics and thermal properties controlled by instantaneous temperature and state of cure, e.g. benzoxazines, bismaleimides and vinyl esters. Matrix systems that involve additional state variables as for example the strong pressure dependence of the cure of a phenolic system, would not be covered by the proposed framework. This is also likely to be the case for thermosetting matrices with phase separating thermoplastic tougheners, in which thermal history might need to be fully represented in terms of material microstructure by the instantaneous state of cure. The development of the framework for continuous carbon fibre materials has a direct generalisation to other continuous fibre reinforcement options, such as glass or basalt. In cases where the architecture of reinforcement involves 3-D characteristic, e.g. 3-D woven materials or through thickness reinforce architecture, the associated framework based on one-dimensional heat transfer would have limitations necessitating more extended datasets and the corresponding modification of the ML models.

Application in a real-time context in the manufacturing environment requires full consideration of uncertainty and variability. A part of this is addressed through consideration of material and boundary condition variability in the current framework. In addition to these, a full coverage of uncertainties needs to incorporate sensor errors and/or model errors where these are applicable. The issue of uncertainty and its incorporation in the framework is accentuated by utilisation within an optimisation framework. Optimisation algorithms exploit landscapes often finding efficient solutions in high gradient/sensitivity areas. In addition, the present validation results — narrow simulation confidence intervals and small experimental deviations within established composite process variation limits — can support initial evidence of reliability for the proof-of-concept scope. However, full industrial adoption will require extending this validation under plant-specific variability, expanded geometrical configurations, and operational constraints. This is further motivated by the fact that an optimal and performance efficient process setup is often also less robust; therefore, integration of variability is of high importance.

From an ML perspective, scalability and generalisation remain tied to the diversity and representativeness of the FE-generated training data. Predictive accuracy is generally highest within the training domain and may degrade for substantially different scenarios. Whilst the current metamodels show narrow confidence intervals and strong agreement with experimental results for the studied case, reliability may decrease for geometries, materials, or process conditions under-represented in the training set. Real-

time deployment also requires careful resource management for inference when scaling to larger datasets or more complex models. Addressing these limitations will require ongoing dataset expansion, integration of interpretability tools, and uncertainty quantification tailored to the intended operational domain.

4. Conclusions

In this study, a machine-learning-based surrogate simulation approach was developed to predict rapidly critical curing process outputs targeting real-time deployment. Initially trained on FE-generated data from a 12 mm RTM plate, the ML models showed excellent convergence and robustness, achieving R^2 values above 0.98 for temperature overshoot and cure time, while reproducing their correlation and Pareto behaviour. The model enables demanding optimisation tasks to be executed within minutes. In addition, to validate the robustness and practical applicability of the models, the ML models were retrained on FE data from a 3 mm thick part and validated through RTM manufacturing trials, with DSC confirming cure completion. Given the narrow simulation confidence intervals and the observed trial deviations well within established process variation limits, the probability of a large undetected discrepancy is low. Future work will expand the experimental dataset to reinforce statistical generalisation and quantify experimental repeatability through statistical replication.

This development enables iterative tasks such as optimisation and stochastic simulation to be executed in real-time during the cure of parts using RTM. Furthermore, the ML models capture correctly the correlation structure of variables and output allowing utilisation within a multi-objective optimisation setting. Potential integration of the ML approach and its predictions with process monitoring data in the future can deliver significant enhancement of automation capabilities. Moreover, the ML models developed in this study can potentially be extended to the injection process, describing the full chain of manufacturing using RTM. The principles of the implementation presented here are identical for processing using other manufacturing methods, such as autoclave or oven curing, with differences mostly around the definition of boundary conditions in the FE models. The end product of this development targets application to composite process design as an off-line tool based on simulation and utilisation of the ML approach as a metamodel as well as real-time application alongside process monitoring to enable a multi-objective setting to be used as the basis of active control to deal with process and material uncertainty. The former is an ideal solution for mass production applications (e.g. automotive parts), whereas the latter is suitable for low volume high value part production such as in aerospace and offshore energy applications. As a future

direction of this work, experimental implementation of the models could enable their deployment in real-time manufacturing process control. Furthermore, advanced ML techniques such as continuous learning and theory-guided ML can be employed to incorporate new simulation results and to reduce the number of required training data points. These process enhancements can enable the real-time ML assisted optimisation and monitoring. The utilisation of ML as a metamodel in this work demonstrates its capabilities and efficiency in dealing with the complexity of the cure problem. However, any shortcomings of the predictive approach based on FE modelling of the physical reality are carried over to the ML solution. This makes the approach vulnerable to potential discrepancies arising due to the complexity of the manufacturing environment not captured by the numerical FE solution. Application of the demonstrated ML using real production data in repetitive manufacturing scenarios will overcome this challenge and guarantee overall validity of the approach.

Declaration of conflicting interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iD

Evrydiki Kyriazi: 0000-0001-9228-4853

Charalampos Zervos: 0000-0003-3162-3561

Georgios Syriopoulos: 0009-0003-6500-1661

Geoffrey Neale: 0000-0003-3731-1059

Mehdi Asareh: 0000-0002-6199-1341

Swaroop Narayanan: 0000-0002-9267-7283

Hercules Avramopoulos: 0000-0001-8793-2262

Alex Skordos: 0000-0003-1273-029X

Author Contributions

Conceptualisation: Kyriazi E., Petsinis G., and Skordos A.;

Methodology: Kyriazi E., Petsinis G., Neale G., Asareh M., and Skordos A.;

Validation: Kyriazi E., Zervos C., Pouloupoulos I., Syriopoulos G., Narayanan S., and Skordos A.;

Formal analysis: Kyriazi E., Neale G., Asareh M., and Skordos A.;

Resources: Lewis S., Avramopoulos H., and Skordos A.;

Data curation: Kyriazi E. and Skordos A.;

Writing—original draft preparation: Kyriazi E., Petsinis G., and Skordos A.;

Writing—review and editing: Kyriazi E., Zervos C., Narayanan S., Page C., and Skordos A.;

Visualisation: Kyriazi E. and Skordos A.;

Supervision: Avramopoulos H. and Skordos A.;

Project administration: Avramopoulos H. and Skordos A.;

Funding acquisition: Avramopoulos H. and Skordos A.

Data availability statement

The simulation datasets and machine learning scripts used in this study are available from the corresponding author upon request.

Acknowledgements

This work has received funding from the European Union’s Horizon 2020 innovation programme under grant agreement No. 871875 (SEER).

References

- [1] N. Rai, R. Pitchumani, Optimal cure cycles for the fabrication of thermosetting-matrix composites, *Polym. Compos.* 18 (1997) 566–581. <https://doi.org/10.1002/pc.10309>.
- [2] P. Carlone, G.S. Palazzo, A Simulation Based Metaheuristic Optimization of the Thermal Cure Cycle of Carbon-Epoxy Composite Laminates, *AIP Conf. Proc.* (2011), <https://doi.org/10.1063/1.3589483>.
- [3] A. Mawardi, R. Pitchumani, Optimal Temperature and Current Cycles for Curing of Composites Using Embedded Resistive Heating Elements, *J. Heat Transf.* 125 (2003) 126–136. <https://doi.org/10.1115/1.1527903>.
- [4] A. Skordos, I. Partridge, Inverse heat transfer for optimization and on-line thermal properties estimation in composites curing, *Inverse Probl. Sci. Eng.* (2004) 157–172. <https://doi.org/10.1080/10682760310001598616>.
- [5] D. Aleksendrić, P. Carlone, Č. Velimir, Optimization of the Temperature-Time Curve for the Curing Process of Thermoset Matrix Composites, *Appl. Compos. Mater.* 23 (2016).

- <https://doi.org/10.1007/s10443-016-9499-y>.
- [6] G. Struzziero, J.J.E. Teuwen, Effect of convection coefficient and thickness on optimal cure cycles for the manufacturing of wind turbine components using VARTM, *Compos. Part Appl. Sci. Manuf.* 123 (2019) 25–36. <https://doi.org/10.1016/j.compositesa.2019.04.024>.
 - [7] R. Matsuzaki, R. Yokoyama, T. Kobara, T. Tachikawa, Multi-objective curing optimization of carbon fiber composite materials using data assimilation and localized heating, *Compos. Part Appl. Sci. Manuf.* 119 (2019) 61–72. <https://doi.org/10.1016/j.compositesa.2019.01.021>.
 - [8] K.I. Tifkitis, T.S. Mesogitis, G. Struzziero, A. Skordos, Stochastic multi-objective optimisation of the cure process of thick laminates, *Compos. Part Appl. Sci. Manuf.* 112 (2018) 383–394. <https://doi.org/10.1016/j.compositesa.2018.06.015>.
 - [9] S. Cassola, M. Duhovic, T. Schmidt, D. May, Machine learning for polymer composites process simulation – a review, *Compos. Part B Eng.* 246 (2022) 110208. <https://doi.org/10.1016/j.compositesb.2022.110208>.
 - [10] J. Mendikute, J. Plazaola, M. Baskaran, E. Zugasti, L. Aretxabaleta, J. Aurrekoetxea, Impregnation quality diagnosis in Resin Transfer Moulding by machine learning, *Compos. Part B Eng.* 221 (2021) 108973. <https://doi.org/10.1016/j.compositesb.2021.108973>.
 - [11] S. Stieber, N. Schröter, A. Schiendorfer, A. Hoffmann, W. Reif, FlowFrontNet: Improving Carbon Composite Manufacturing with CNNs, in: Y. Dong, D. Mladenović, C. Saunders (Eds.), *Mach. Learn. Knowl. Discov. Databases Appl. Data Sci. Track*, Springer International Publishing, Cham, 2021: pp. 411–426. https://doi.org/10.1007/978-3-030-67667-4_25.
 - [12] T. Lavaggi, M. Samizadeh, N. Niknafs Kermani, M.M. Khalili, S.G. Advani, Theory-guided machine learning for optimal autoclave co-curing of sandwich composite structures, *Polym. Compos.* 43 (2022) 5319–5331. <https://doi.org/10.1002/pc.26829>.
 - [13] N. Zobeiry, D. Ee, Floyd, A. Poursartip, Theory-Guided Machine Learning Composites Processing Modelling for Manufacturability Assessment in Preliminary Design, in: 2019 <https://doi.org/10.48550/arXiv.2103.16010>
 - [14] Y. Wang, S. Xu, K.H. Bwar, B. Eisenbart, G. Lu, A. Belaadi, B. Fox, B.X. Chai, Application of machine learning for composite moulding process modelling, *Compos. Commun.* 48 (2024) 101960. <https://doi.org/10.1016/j.coco.2024.101960>.
 - [15] A. Mujtaba, F. Islam, P. Kaeding, T. Lindemann, B. Gangadhara Prusty, Machine-learning based process monitoring for automated composites manufacturing, *J. Intell. Manuf.* 36 (2025) 1095–1110. <https://doi.org/10.1007/s10845-023-02282-2>.
 - [16] C. Wu, R. Zhang, P. Zhao, L. Li, D. Zhang, Curing simulation and data-driven curing curve prediction of thermoset composites, *Sci. Rep.* 14 (2024) 31860. <https://doi.org/10.1038/s41598-024-83379-3>.
 - [17] F. Zhang, H. Wang, Y. Liu, H. Guo, W. Wang, J. Du, Fast prediction of curing residual stress in polymer-matrix composites by a deep learning method, *Polym. Compos.* 46 (2025). <https://doi.org/10.1002/pc.29501>.
 - [18] K.D. Humfeld, G.Y. Kim, J.H. Jeon, J. Hoffman, A. Brown, J. Colton, S. Melkote, V. Nguyen, Co-training of multiple neural networks for simultaneous optimization and training of physics-informed neural networks for composite curing, *Compos. Part Appl. Sci. Manuf.* 193 (2025) 108820. <https://doi.org/10.1016/j.compositesa.2025.108820>.
 - [19] K.I. Tifkitis, A. Winistoerfer, A. Skordos, Online optimisation and active control of the cure process of thick composite laminates, *J. Manuf. Process.* 87 (2023) 221–230. <https://doi.org/10.1016/j.jmapro.2023.01.027>.
 - [20] P. Karkanias, I. Partridge, Cure modeling and monitoring of epoxy/amine resin systems. I. Cure kinetics modeling, *J. Appl. Polym. Sci.* 77 (2000) 1419–1431. [https://doi.org/10.1002/1097-4628\(20000815\)77:7<1419::AID-APP3>3.0.CO;2-N](https://doi.org/10.1002/1097-4628(20000815)77:7<1419::AID-APP3>3.0.CO;2-N).
 - [21] J.P. Pascault, R.J.J. Williams, Relationships between glass transition temperature and conversion, *Polym. Bull.* 24 (1990) 115–121. <https://doi.org/10.1007/BF00298330>.
 - [22] T. Mesogitis, A. Skordos, A.C. Long, Stochastic simulation of the influence of cure kinetics uncertainty on composites cure, (2015). <https://doi.org/10.1016/j.compscitech.2015.02.009>.
 - [23] G. Struzziero, A. Skordos, Multi-objective optimisation of the cure of thick components, *Compos. Part Appl. Sci. Manuf.* 93 (2017) 126–136. <https://doi.org/10.1016/j.compositesa.2016.11.014>.
 - [24] J.D. Farmer, E.E. Covert, Thermal conductivity of a thermosetting advanced composite during its cure, *J. Thermophys. Heat Transf.* 10 (1996) 467–475. <https://doi.org/10.2514/3.812>.
 - [25] Marc® 2023.2 Volume A: Theory and User Information. Hexagon, 2023
 - [26] Marc® 2023.2 Volume B: Element library. Hexagon; 2023

- [27] Marc® 2023.2 Volume D: User subroutines and Special Routines. Hexagon; 2023
- [28] A. Skordos, I. Partridge, Cure kinetics modeling of epoxy resins using a non-parametric numerical procedure, *Polym. Eng. Sci.* 41 (2001) 793–805. <https://doi.org/10.1002/pen.10777>.
- [29] Hexcel® RTM 6 180 °C epoxy system for Resin Transfer Moulding monocomponent system Product Data. www.hexcel.com; 2009
- [30] T. Mesogitis, A. Skordos, A. Long, Stochastic heat transfer simulation of the cure of advanced composites, *J. Compos. Mater.* 50 (2016) 2971–2986. <https://doi.org/10.1177/0021998315615200>.
- [31] Hexcel® RTM 6 180 °C epoxy system for Resin Transfer Moulding monocomponent system Product Data. www.hexcel.com; 2016.
- [32] T. Jiang, J.L. Gradus, A.J. Rosellini, Supervised Machine Learning: A Brief Primer, *Behav. Ther.* 51 (2020) 675–687. <https://doi.org/10.1016/j.beth.2020.05.002>.
- [33] F. García Fernández, P. de Palacios, A. García-Iruela, L.G. Esteban, Using Bootstrapping to Determine Artificial Neural Network Confidence Intervals—Case Study of Particleboard Internal Bond Determined from Production Data, *Appl. Sci.* 15 (2025) 4554. <https://doi.org/10.3390/app15084554>.
- [34] J.H. Friedman, Greedy function approximation: A gradient boosting machine., *Ann. Stat.* 29 (2001) 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
- [35] T.O. Omotehinwa, D.O. Oyewola, Hyperparameter Optimization of Ensemble Models for Spam Email Detection, *Appl. Sci.* 13 (2023) 1971. <https://doi.org/10.3390/app13031971>.
- [36] B. Wang, J. Zhao, Z. Guo, B. Wang, Failure envelope prediction of 2D SiCf/SiC composites based on XGBoost model, *Compos. Part Appl. Sci. Manuf.* 185 (2024) 108287. <https://doi.org/10.1016/j.compositesa.2024.108287>.
- [37] L. Zhang, Z. Guo, Q. Tao, Z. Xiong, J. Ye, XGBoost-based short-term prediction method for power system inertia and its interpretability, *Energy Rep.* 9 (2023) 1458–1469. <https://doi.org/10.1016/j.egyr.2023.04.065>.
- [38] M. Wortsman, G. Ilharco, S.Y. Gadre, R. Roelofs, R. Gontijo-Lopes, A.S. Morcos, H. Namkoong, A. Farhadi, Y. Carmon, S. Kornblith, L. Schmidt, Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time, in: *Proc. 39th Int. Conf. Mach. Learn.*, PMLR, 2022: pp. 23965–23998. <https://proceedings.mlr.press/v162/wortsman22a.html>.
- [39] M. Shahhosseini, G. Hu, H. Pham, Optimizing ensemble weights and hyperparameters of machine learning models for regression problems, *Mach. Learn. Appl.* 7 (2022) 100251. <https://doi.org/10.1016/j.mlwa.2022.100251>.
- [40] J. Lee, M. Duhovic, D. May, T. Allen, P. Kelly, Physics-informed neural networks for real-time simulation of transverse Liquid Composite Moulding processes and permeability measurements, *Compos. Part Appl. Sci. Manuf.* 193 (2025) 108857. <https://doi.org/10.1016/j.compositesa.2025.108857>.
- [41] Mulligan T, Gnaniah S, Sims G. Measurement Good Practice Guide No. 62: Thermal analysis techniques for composites and adhesives, 2nd Edition, National Physical Laboratory
- [42] TRIOS Software, Waters™, <https://www.tainstruments.com/trios-software/>.
- [43] T. Mesogitis, J. Kratz, A. Skordos, Heat transfer simulation of the cure of thermoplastic particle interleaf carbon fibre epoxy prepregs, *J. Compos. Mater.* 53 (2019) 2053–2064. <https://doi.org/10.1177/0021998318818245>.
- [44] Y. Guo, B. Yang, S. Alqawzai, K. Chen, D. Kong, Machine learning-based and interpretable models for predicting the resistance and probability of progressive collapse of steel frame-composite floor structures, *Eng. Struct.* 332 (2025) 120089. <https://doi.org/10.1016/j.engstruct.2025.120089>.



DSPACE

<https://dspace.org/>

ML-based surrogate cure simulation for predicting process time and temperature overshoot in resin transfer moulding

Kyriazi, Evrydiki; Petsinis, Georgios; Zervos, Charalampos; Pouloupoulos, Ioannis; Syriopoulos, Georgios ...

2025-12-31

Sage

<https://doi.org/10.1177/07316844251390926>

Kyriazi E, Petsinis G, Zervos C, et al., (2025) ML-based surrogate cure simulation for predicting process time and temperature overshoot in resin transfer moulding. Journal of Reinforced Plastics and Composites, Available online 4 November 2025