Theory-Guided Machine Learning Composites Processing Modelling for Manufacturability Assessment in Preliminary Design

Conference Paper · June 2019

CITATIONS
7

READS
1,139

4 authors, including:

Navid Zobeiry
University of Washington Seattle
95 PUBLICATIONS **1,033** CITATIONS

SEE PROFILE

A. Poursartip
University of British Columbia - Vancouver
179 PUBLICATIONS **4,210** CITATIONS

SEE PROFILE

Theory-Guided Machine Learning Composites Processing Modelling for Manufacturability Assessment in Preliminary Design

Dr. N. Zobeiry
(The University of British Columbia, Canada)

D. van Ee, Dr. A. Floyd
(Convergent Manufacturing Technologies, Canada)

Dr. A. Poursartip
(Convergent Manufacturing Technologies, Canada
and The University of British Columbia, Canada)

Abstract

Science-based simulation tools, where physical laws are explicitly coded and solved using numerical techniques such as finite element analysis, are widely used. Recent advances in artificial intelligence and machine learning algorithms as well as the emergence of "big data" have resulted in the development of data-driven models where domain knowledge is not necessarily required. Unfortunately, these "theory-agnostic black-box data science" models have achieved limited success in engineering applications. Theory-guided machine learning (TGML) is an emerging approach that integrates domain knowledge with machine learning. To introduce physical consistency in training machine learning models, this approach incorporates the physics of the problem using proxies such as physics-based features, model architecture, activation functions, loss functions, and constrained response surfaces to achieve higher-fidelity models while requiring less training data.

Composites process modelling can be a computationally expensive task. Typical process simulations are multi-scale and multi-physics in nature and require the simulation of real-world processes that occur over many hours. For example, a typical process simulation of the curing of a composite part might involve a thermal analysis of a part, its bagging material, its tooling, and the tooling support structure in an oven or autoclave. After the thermal simulation, which calculates the temperature and material properties everywhere, a saturated flow analysis would be performed, capturing the resin flow and fibre movement during the curing process. Then, building on the thermal and flow analyses, a mechanical simulation would be performed to determine the ultimate part deformation (spring-in) and residual stress state. Serially-coupled analyses of parts at an industrially

relevant scale may take several to many days to complete, even on HPC clusters. The ability of engineers to run multiple simulations to explore the design space and create optimized process parameters is severely limited by this computational expense.

TGML models can be trained to simulate simplified process models. For example, the manufacture of a single-layer composite part can be represented by 8 input parameters. In the case of thermal analyses, process key performance indicators (such as maximum exotherm temperature, minimum viscosity, etc.) may be identified. Using TGML, it is demonstrated that a small and manageable number of training simulations can be used to train a model that accurately captures the behaviour.

With TGML models that simulate aspects of the full composites processing simulation chain, numerically efficient tools can be used to evaluate designs for manufacturability, even at the preliminary design stage. These tools take advantage of the speed and accuracy of the TGML models to quickly perform many analyses that fully explore the response of a given part design given typical process parameters in order to assess the likelihood that the part, as designed, can be manufactured within the required specifications. It is demonstrated that the speed-up of such tools resulting from the use of TGML models rather than finite element simulations makes the tools practical for use in industrial practice.

1. Introduction

With the growing ability to measure, generate, and store very large sets of data, and with a corresponding growth in the ability to perform computations very, very quickly, machine learning has experienced an incredible growth in recent years [1]. Inspired by how learning occurs in organisms, machine learning is a term used to describe a collection of algorithms that are used to generate complex models from a large set of inputs with limited guidance. The machine-learning system "teaches itself" how to solve a complex problem.

This, of course, is an over-simplification. An incredible amount of research has gone into different aspects of machine learning including: the algorithms used to link the inputs to the outputs, choosing the appropriate types of data, choosing an appropriate number of data, evaluating the performance of the models, training the models, and so on. The system may "teach itself", but an astonishing level of effort typically goes into preparing all aspects of a system before it can be used.

Machine-learning tends to be multi-disciplinary in nature; in addition to computer science and data science, domain experts are involved, and machine-learning has

been applied to an extremely wide range of topics. Machine learning is used in applications such as voice and image recognition, automated driving, spam email filtering, machine failure prediction in manufacturing, and so on. Despite these successes, use of machine learning in engineering domains, particularly as a predictor for “hard-science” outcomes, has been limited. In these engineering applications, data tends to be small in number, fragmented, and suffer from “underrepresented classes” – cases where the number of “successes” vastly outnumber the number of “failures” – so predicting failure becomes very difficult [2].

Theory-guided machine learning is a term used to describe incorporating domain-based scientific theory and phenomenological models into the machine learning algorithms [2, 3]. Essentially, the ML models are being intentionally biased to known domain-specific behaviours. Introducing theory into the machine learning process can vastly reduce the amount of data required to successfully train a ML model, and improve the reliability of model predictions outside of the training data.

2. Composites Process Modelling for Manufacturability Assessment in Preliminary Design

The manufacture of parts made from composite materials is the last step in a long workflow that incorporates various types of design from engineers with different specialities. For example, a aerospace structural design is created from load and shape requirements (aero surfaces). The materials used in the structural design likely have been down-selected by a material and process engineer. A tooling engineer will create the design of the substructure that supports the part being made. A facilities engineer has designed the equipment that will be used to cure the part (an autoclave, for example). Traditionally, the down-stream effects of design decisions are not well understood at up-stream levels.

As an example, consider the design of a composite wing skin. A representative wing skin design is shown in Figure 1. This notional wing skin is full-size, approximately 10 m x 5 m in size, with a representative ply definition consisting of pad-ups, ply drops, and other features present in a typical composite wing skin. The skin thickness varies from 40 to 80 plies.

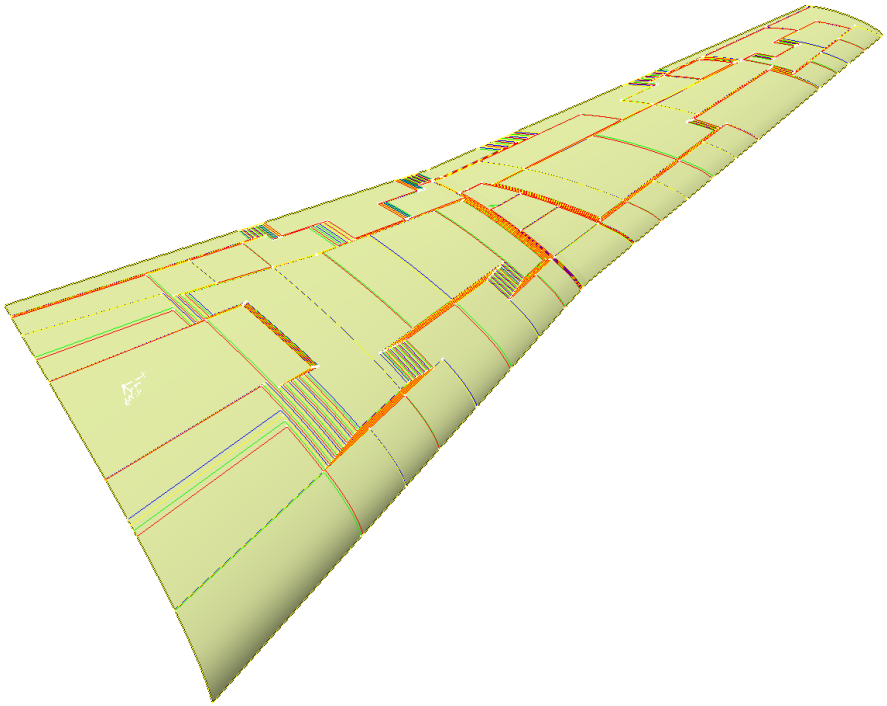

Figure 1: Representative wing skin

During the preliminary design phase of a part like this, it is possible to make decisions that inadvertently negatively affect the manufacturability of the part. For example, the net cost of dealing with a specification of a certain layup may be quite high if that layup causes an exotherm in the cure process using typical tooling and cure processes. Unfortunately, the designer typically does not have the processing expertise to anticipate the consequences that may arise during manufacturing. This can lead to longer process development timelines and increased cost bringing products to market.

If an expert materials and process analyst were to investigate the thermal behaviour of the wing skin using process modelling, the workflow would be:

1. Convert the geometry to a grid, based on spar and rib locations
2. Create zones of similar layup based on the grid
3. Mesh the surface (in this case, ~10,000 elements)
4. Export mesh to FE package (e.g., ABAQUS)
5. Extrude shell mesh to solid mesh
6. Define tooling geometry, mesh, transfer to FE package
7. Set thermal analysis parameters:
    a. Materials (composites, tooling, bagging, breather, etc)

   b. Cure cycle (air temperature load)
   c. Heat transfer coefficients at each surface
8. Perform thermal analysis
9. Evaluate simulation for:
   a. Excessive temperatures (exotherms)
   b. Thermal lags during cure
   c. Material properties during cure and at end of cure cycle

Convergent has developed a design aid to make this assessment of manufacturability possible for engineers without a background in composites manufacturing, in an automated and less time-intensive workflow. The software product, Composites Producibility Assessment – Thermal Analysis (CPA-TA), examines a given design and performs an assessment of the part details using typical processing conditions for the particular factory. This assessment is completed by performing process simulations of one dimensional “drill-throughs” of each unique geometric section of the part using typical values for heat transfer coefficient, cure cycles, tooling materials and thicknesses, and so on. Figure 2 shows an example of using CPA-TA with this representative wing skin in CATIA V5.

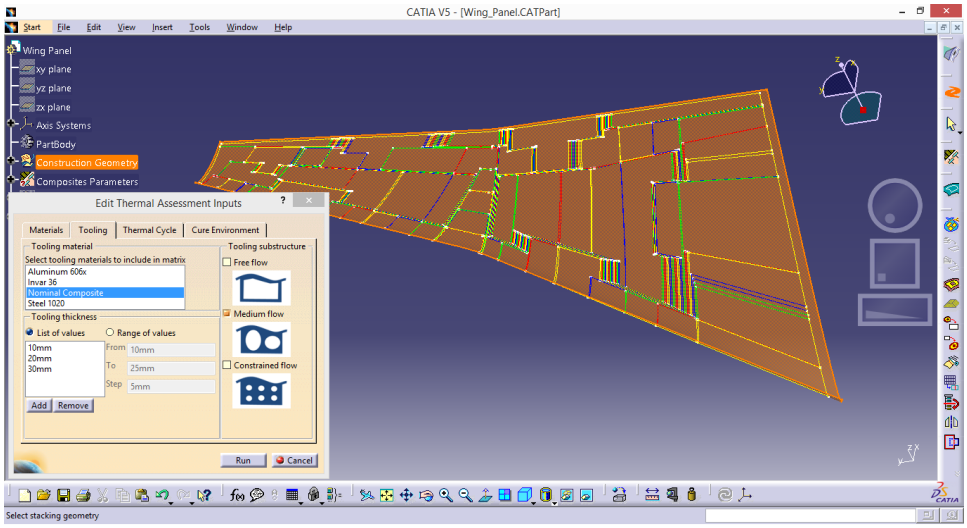

Figure 2: Using CPA-TA in CATIA to evaluate the manufacturability of a composite wing skin

Evaluating the whole process space is a computationally expensive exercise and can take quite some time to complete. For example, the representative wing skin has 63 zones, in which there are 30 unique 1D geometries (see Figure 3). Evaluating only three different cure cycles and three tool face sheet thicknesses

requires 270 unique analyses. On a typical engineering workstation, these analyses took a total of 1620 seconds (27 minutes) to complete.

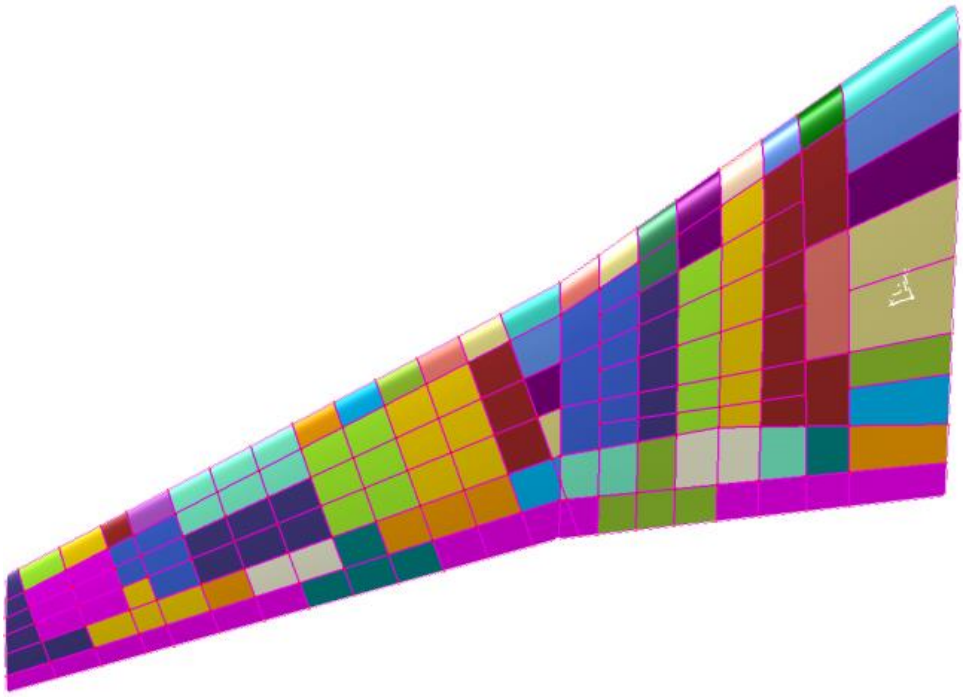

Figure 3: Zones of unique 1D geometry

To reduce the time spent waiting for the computer to evaluate the design, different strategies can be considered. For example, instead of examining each unique geometry, thicknesses may be rounded to a convenient value (nearest 1 mm or 10 mm for example). Using this approach reduced the number of required runs to 45 (from 270) and the time to perform all the analyses to 276 seconds (about 4.6 minutes). The trade off is a loss of accuracy.

If these simulations have been performed in the past (the same geometry and analysis conditions), the results can be stored in a cache. Performing the analyses for the rounded geometries after having run them once already reduced the computation time to 38 seconds.

Rather than perform the analyses directly on the designer's workstation, or having to rely on the designer having run the analyses previously, theory-guided machine learning can used as a surrogate engine to replace the process simulation analyses.

3. Machine Learning Concepts

Before explaining the application of TGML to process simulation of composite materials, a few machine-learning concepts must be introduced.

A typical machine-learning algorithm that forms the core of the ML “engine”, is the artificial neural network (ANN). An artificial neural network is that algorithm that links the set of inputs to the output, by using the concept of layers of artificial neurons (see [4] for example). The neurons are represented by discrete values in the algorithm, either the inputs or outputs themselves, or intermediate values used in the overall calculation of the outputs.

These neurons are conceptually arranged in layers. The set of inputs forms the input layer, and the set of outputs forms the output layer. In between the input and output layer are a number of intermediate layers (or “hidden layers”) that are made up of a number of neurons. The number of hidden layers and number of neurons that make up each of those hidden layers is a key aspect to developing a successful and efficient artificial neural network system.

The neurons in each layer are connected to neurons in neighbouring layers by “edges”. These edges represent the contribution of each neuron to the calculation of the state of the next neuron. Determining the weighting of these contributions of connected neurons is a key aspect of training the neural network.

A simplified ANN with three input values and two output values is shown in Figure 4. This ANN has a single hidden layer with four neurons. The edges are shown in the figure, illustrating the relationship between neurons of neighbouring layers.

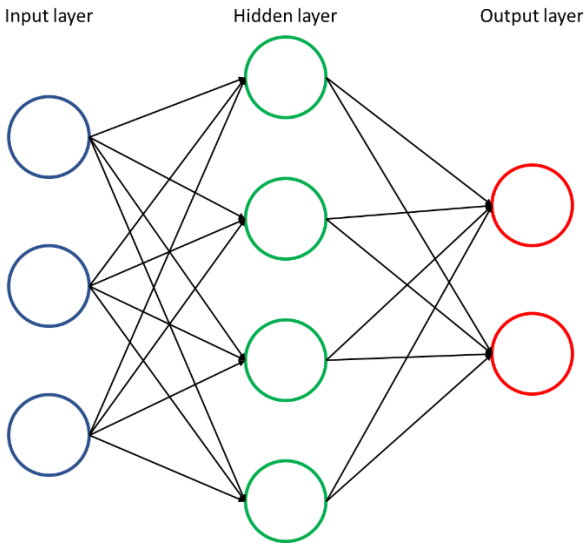

Figure 4: Idealized representation of an artificial neural network (ANN).

Each artificial neuron uses a particular function to transform its weighted inputs into an output value. This function is called an "activation function". The selection of the activation function for each neuron or layer of neurons is a crucial step in the design of an ANN. The choice of activation function should be dependent on the relationship between the inputs and the outputs. Some "general purpose" activation functions are often used, but domain-knowledge can guide the selection of the activation function.

Training the neural network requires a large set of data. This data is typically divided into two subsets: one subset used to train the network (the "training data"), and the other used to determine how well the ANN can predict the data set (the "validation data"). When training the ANN, the training algorithm is often run through many iterations. How well the ANN predicts the data (both the training data during training, and the validation data after training) can be measured a number of different ways. A very simple way of evaluating the goodness-of-fit is to measure the root-mean-squared-error (RMSE) and maximum error. The RMSE is calculated by taking the square root of the mean of the squares of residual error for each point:

$$RMSE = \sqrt{\frac{\sum_1^N (f(x_i) - y_i)^2}{N}}$$

The maximum error is the normalized maximum difference between the observed and calculated value over the whole data set.

Training the ANN requires the selection of an algorithm to perform the training. Many algorithms exists, and the selection of an appropriate algorithm, its parameters, and the number of iterations to run the training are all part of creating a successful ANN model. In this work, a “proximal” optimizing algorithm was used throughout.

4. TGML in Thermal Process Simulation of Composite Materials

To introduce theory-guided machine learning into process simulation, we must idealize the system to a set of well-defined inputs and outputs. The one-dimensional drill-through simulations can be parameterized with a small number of inputs. Each drill-though consists of a single layer of part material of thickness L_1 and an initial degree of cure x_i, a single layer of inert tool material of thickness L_2, a heat transfer coefficient on the top and bottom of the stack (h_1, h_2), and a heating cycle. This is illustrated in Figure 5.

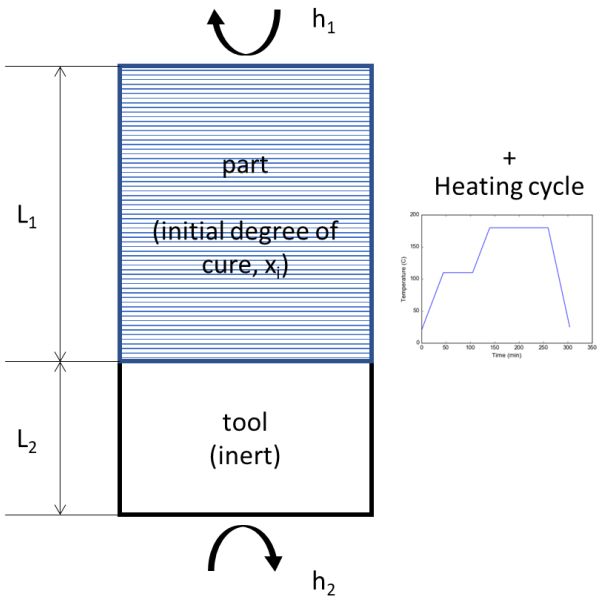

Figure 5: Parameterized 1-D drill-through.

The performance of the 1-D stack can be assessed by observing key performance indicators (KPIs) and comparing them to desired outcomes or limitations. In this example, the KPIs of interest are:

1. Exotherm: the maximum overshoot of the part temperature above the air temperature over the heating cycle
2. Thermal lag: the maximum lag of the part temperature under the air temperature over the heating cycle
3. Minimum viscosity: the minimum viscosity observed in the resin of the part material over the heating cycle.

These KPIs are illustrated in Figure 6.

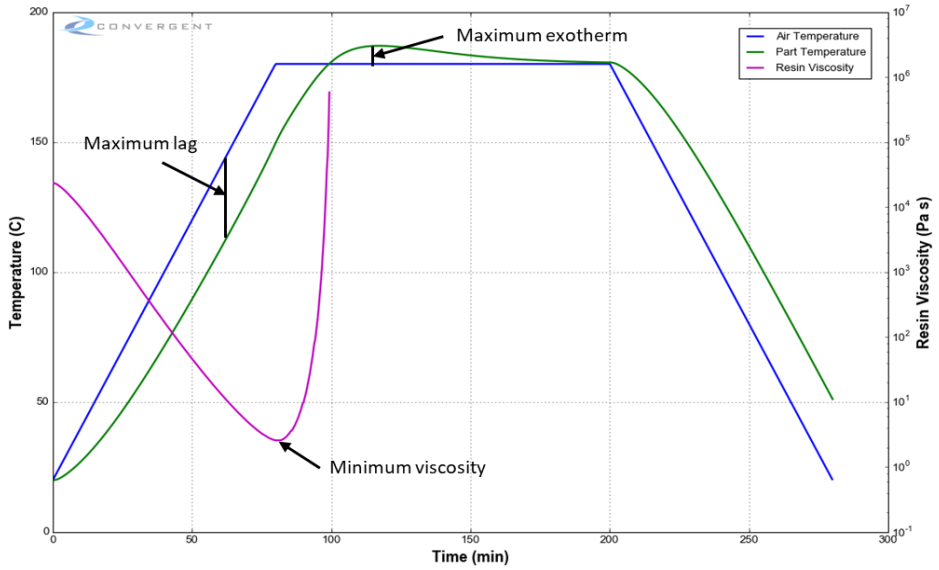

Figure 6: Simulations results from a typical 1D analysis illustrating the KPIs of interest.

The reduction of the problem to a set of inputs and outputs makes it well-suited for machine learning. A large data set of 1D simulations can be generated and used to create a neural network. The resulting neural network can then be used as a substitute for the simulation (a “surrogate engine”), with the ability to very quickly generate accurate predictions without having to perform the complete process simulation.

In this case, however, the processing science of this idealized drill-though is well understood. Rather than creating a typical neural network, the domain-specific scientific understanding of the phenomena involved in the manufacturing process can be leveraged to create a better machine-learning model that requires less training data and performs better outside of the range of the training data.

A corpus of training data was created using the RAVEN process modelling software [5]. An AS4/8552 material file set was chosen for use in this investigation, based on an open-literature characterization [6]. This material has a manufacturer’s recommended cure cycle that is a two-hold cycle: ramp to 110 °C and hold for 60 minutes, then ramp to 180 °C and hold for 120 minutes. Figure 7 shows a typical heating cycle for this material.

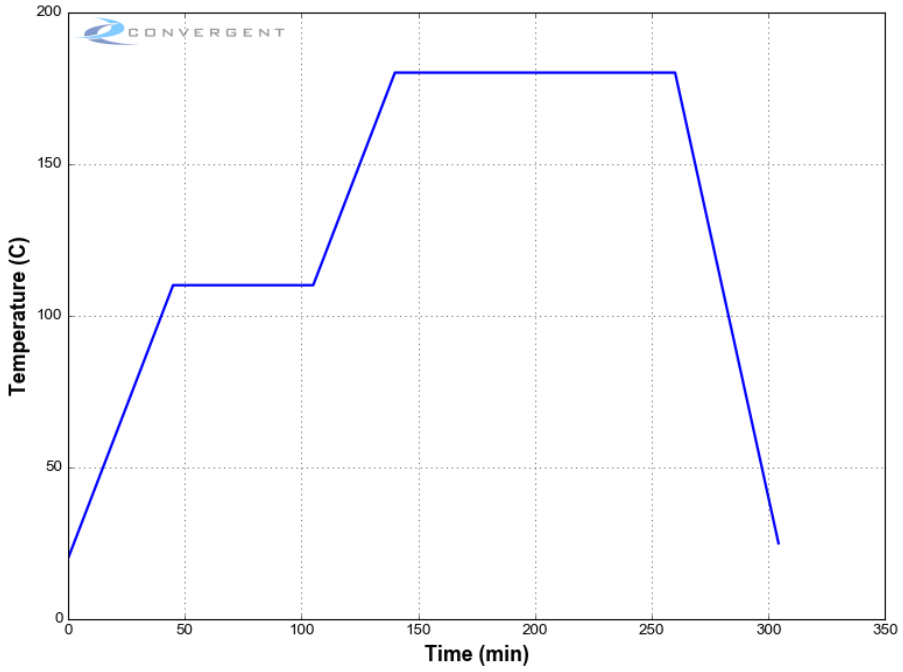

Figure 7: Typical cure cycle for AS4/8552.

20,000 1D simulations were performed for each of three different tooling materials based on the following simulation matrix. Note that data points were taken with a random distribution from within the parameter variations to avoid biasing the data.

Table 1: Distribution of parameters for RAVEN analyses.

Parameter	Inputs
Part material	AS4/8552
Part thickness (L1)	2 mm – 20 mm
Tool material	Invar, steel, composite

Tool thickness (L2)	8 mm – 20 mm
Top HTC (h1)	20 W/(m^2 K) – 100 W/(m^2 K)
Bottom HTC (h2)	20 W/(m^2 K) – 100 W/(m^2 K)
Heating rate	0.5 C/min – 5 C/min

This data took roughly 9 days to generate on a single core.

Application of TGML to each of the three KPIs is considered below.

4.1 Maximum exotherm

To introduce domain knowledge in the prediction of maximum exotherm, we follow the approach of Zobeiry and Poursartip [7] with the following transformed heating rate to predict exotherm:

$$f(T') = \frac{1}{1 + e^{-T'}} - 0.6$$

This equation is obtained based on first principles and understanding the effect of heating rate on exotherm. The constant of 0.6 is obtained through the use of RAVEN simulations. In addition, as discussed by Zobeiry and Poursartip [7], the underlying exponential behaviour of governing equations lead to the use of the *tanh* activation function.

Table 2: TGML models trained to predict maximum exotherm.

Inputs	Tool	Activation Function	Data Points	Optimizer	Hidden Layers	RMSE (°C)	Max Error (°C)
L1, L2, h1, h2, T'	Invar	ReLU	5000	Proximal rate=0.06 n=100000	4×10	0.12	0.6
L1, L2, h1, h2, T'	Invar	tanh	5000	Proximal rate=0.06 n=100000	4×10	0.07	0.7
L1, L2, h1, h2, f(T')	Invar	tanh	5000	Proximal rate=0.06 n=100000	4×10	0.06	0.5

L1, L2, h1, h2, f(T’)	Steel	tanh	5000	Proximal rate=0.06 n=100000	4×10	0.08	0.5
L1, L2, h1, h2, f(T’)	Composite	tanh	5000	Proximal rate=0.06 n=100000	4×10	0.13	1.0

Although all the models predict the maximum exotherm with high accuracy in the training zone, only models that are trained with TGML and the transformed heating rate can predict the trends beyond the boundary of the training zone. This is illustrated in Figure 8, below.

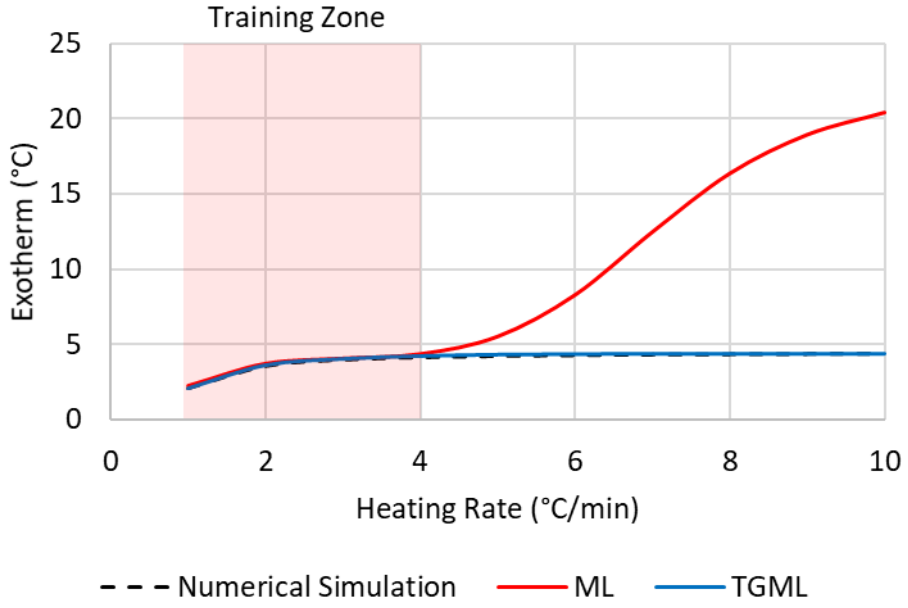

Figure 8: Prediction of maximum exotherm for Invar tooling, L_2=10mm, h_1=60 W/m^{2}K, h_2=40 W/m^{2}K, L_1=8mm

4.3 Maximum Lag

For most cases, maximum lag occurs before entering the hold of 110 °C. This is because negligible curing occurs before the first hold. In the second ramp, exotherm reduces the lag for slow heating rates. Also, in the first ramp, the part is heated from 20 to 110 °C (90 °C difference) while in the second ramp from 110 to 180 °C (70 °C difference). This means that for thermal lag in the current two-hold

cycle, we can neglect cure kinetics and approximately calculate the lag before entering the hold of 110 °C.

For transient thermal lag of a two-layer inert slab, the following approximate equation is known [7, 8, 9, 10]:

$$Transient\ Lag = \Delta T_{SS}\left(1 - e^{-\frac{T_{hold}-T_0}{\Delta T_{SS}}}\right)$$

In which ΔT_{SS} is the steady-state thermal lag and has a linear correlation to heating rate [9, 10]:

$$\Delta T_{SS} = T' \times k(h_1, L_1, L_2, h_2)$$

Where k is a function of other parameters in the system. There are several approaches available to incorporate this causation. Similar to the approach used in the maximum exotherm model, a transformation for heat rate is considered:

$$h(T') = \Delta T_{SS}\left(1 - e^{-\frac{90}{\Delta T_{SS}}}\right)$$

90 °C in above equation is the difference between the hold temperature and room temperature. ANN models were trained based on these transformations:

Table 3: TGML models trained to predict maximum lag.

Inputs	Tool	Activation function	Data points	Optimizer	Hidden layers	RMSE (°C)	Max Error (°C)
L1, L2, h1, h2, T'	Invar	tanh	5000	Proximal rate=0.06 n=200000	5×10	0.2	1.0
L1, L2, h1, h2, h(T')	Invar	tanh	5000	Proximal rate=0.06 n=200000	5×10	0.2	0.9
L1, L2, h1, h2, h(T')	Steel	tanh	5000	Proximal rate=0.06 n=200000	5×10	0.2	1.5

L1, L2, h1, h2, h(T’)	Composite	tanh	5000	Proximal rate=0.06 n=200000	5×10	0.1	0.7

As in the maximum exotherm case, only TGML models can accurately capture the underlying physical trends beyond the boundary of the training zone. This is shown in Figure 9.

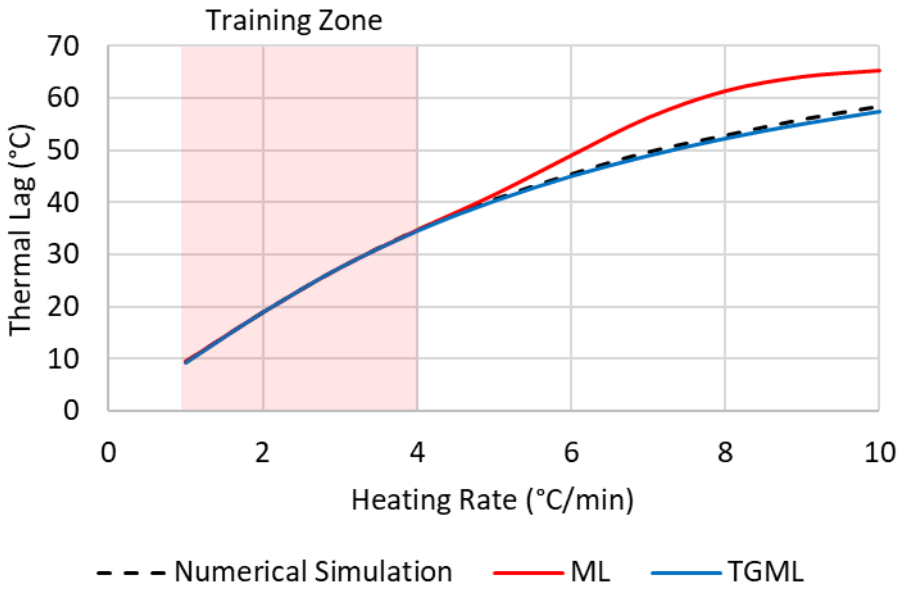

Figure 9: Prediction of thermal lag for Invar tooling, L_2=10mm, h_1=60 W/m^{2}K, h_2=40 W/m^{2}K, L_1=8mm

4.1 Minimum viscosity

Minimum part viscosity occurs well before gelation, so heat generation during polymerization has negligible effect on minimum viscosity. In addition, minimum viscosity occurs at highest temperature locations. Neglecting heat generation, the highest temperature occurs on the surface of the part. Since this is a surface response, thicknesses and heat transfer coefficients have negligible effects on minimum viscosity compared to heating rate. Therefore, the 0D analysis of the problem would be a good approximation for the minimum viscosity behaviour. The following material models are used for the 8552 resin system:

$$\mu = A_1 e^{\frac{E_1}{RT}} \left(\frac{x_{gel}}{x_{gel} - x} \right)^{A+Bx} \cong A_1 e^{\frac{E_1}{RT}} e^{Kx}$$

$$\frac{dx}{dt} = A_2 e^{\frac{-E_2}{RT}} \frac{x^m (1-x)^n}{1 + e^{C(x - x_{C0} - x_{CT} T)}} \cong A_2 e^{\frac{-E_2}{RT}} x^m (1-x)^n$$

During heat-up, temperature is calculated based on the heating rate as follows:

$$T = T_0 + t \times T'$$

Considering the initial degree of cure of 0.05, by the end of the hold temperature at 110 °C, degree of cure is about 0.1 for the range of considered heating rates. For very fast heating rates, negligible curing occurs before reaching the final hold temperature of 180 °C. Therefore, for very fast heating rates, minimum viscosity occurs at the hold temperature of 180 °C with the degree of cure of 0.1. On the other hand, for slower heating rates, curing advances during the second heat-up. Therefore, as temperature is increased, initially viscosity drops. But as curing advances, viscosity increases. Therefore, the minimum viscosity occurs during the second heat-up for slower heating rates. Considering this causation, and also the exponential behaviour of both viscosity and cure as observed in material models, an exponential correlation between minimum viscosity and $1/T'$ can be envisioned. This is confirmed by using RAVEN to predict viscosity at several heating rates:

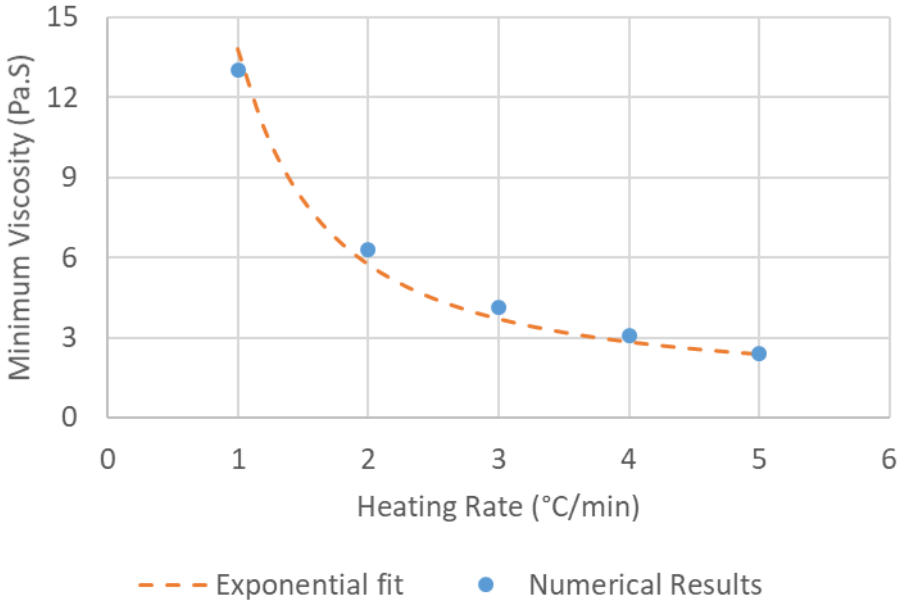

Figure 10: Numerical prediction of minimum viscosity for AS4/8552 using RAVEN as a function of heating rate.

As expected, Figure 10 shows an exponential form for minimum viscosity as a function of heating rate. Using these predictions, the following fitted function can be found:

$$g(T') = e^{\frac{5}{1+T'}}$$

This equation essentially expresses the causation explained earlier. To demonstrate the effect of this causation, artificial neural networks were trained using both traditional ML, and also implementing a transformed heating rate (TGML). The following inputs and hyper-parameters were used to predict minimum viscosity:

Table 4: TGML models trained to predict minimum viscosity.

Inputs	Tool	Activation	Data Points	Optimizer	Hidden Layers	RMSE (Pa s)	Max Error (Pa s)
L1, L2, h1, h2, T’	Invar	tanh	1000	Proximal rate=0.06 n=50000	3×5	0.05	0.2

L1, L2, h1, h2, g(T')	Invar	tanh	1000	Proximal rate=0.06 n=50000	3×5	0.06	0.3
L1, L2, h1, h2, g(T')	Steel	tanh	1000	Proximal rate=0.06 n=50000	3×5	0.06	0.4
L1, L2, h1, h2, g(T')	Composite	tanh	1000	Proximal rate=0.06 n=50000	3×5	0.06	0.4

Once again, only models trained using TGML and the transformed heating rate predict the underlying trends beyond the boundary of the training zone. This is shown in Figure 11 where the results of two models trained with ML and TGML are compared with results from RAVEN simulations.

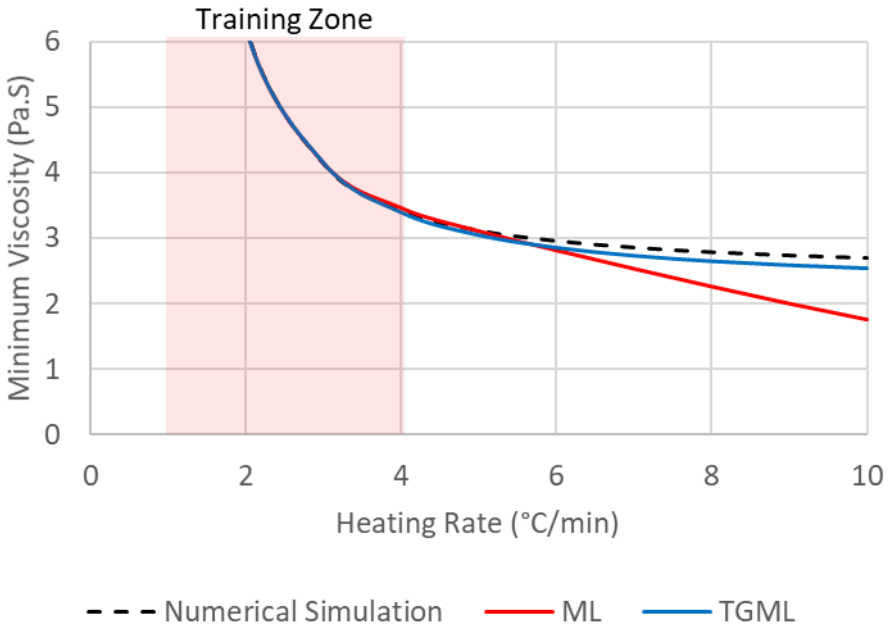

Figure 11: Prediction of minimum viscosity for Invar tooling, L_2=10mm, h_1=60 W/m^{2}K, h_2=40 W/m^{2}K, L_1=8mm.

4.4 TGML Implementation Summary

Along with *tanh* activation functions, following theory-guided transformations were used as inputs for machine learning:

Table 5: Heating rate transformation summary.

KPI	Heating rate transformation
Maximum Exotherm	$f(T') = \frac{1}{1+e^{-T'}} - 0.6$
Thermal Lag	$h(T') = \Delta T_{SS}\left(1 - e^{-\frac{90}{\Delta T_{SS}}}\right)$ $\Delta T_{SS} = T' \times k(h_1, L_1, L_2, h_2)$
Minimum viscosity	$g(T') = e^{\frac{5}{1+T'}}$

4.5 TGML vs ML

To explore the benefits of introducing theory-guided machine learning into composites process simulation, further analyses were performed. Two factors were of particular interest: how many training data points are required to train a good surrogate model, and how does unguided (or “pure”) machine learning compare to TGML with similar inputs.

The effect of both these factors are shown in the figures below. The unguided machine-learning approach used the heating rate (T’) without transformation. Both the unguided and theory-guided machine learning approaches used the *tanh* activation functions, however. Additionally, the unguided and TGML approaches were trained with heating rates between 1 °C/min and 4 °C/min, but then the TGML was also trained on a wider range of data, with heating rates between 0.5 °C/min and 5 °C/min. This expansion of the range was to explore the TGML behaviour, particularly at low heating rates where some interesting modelling behaviour was observed.

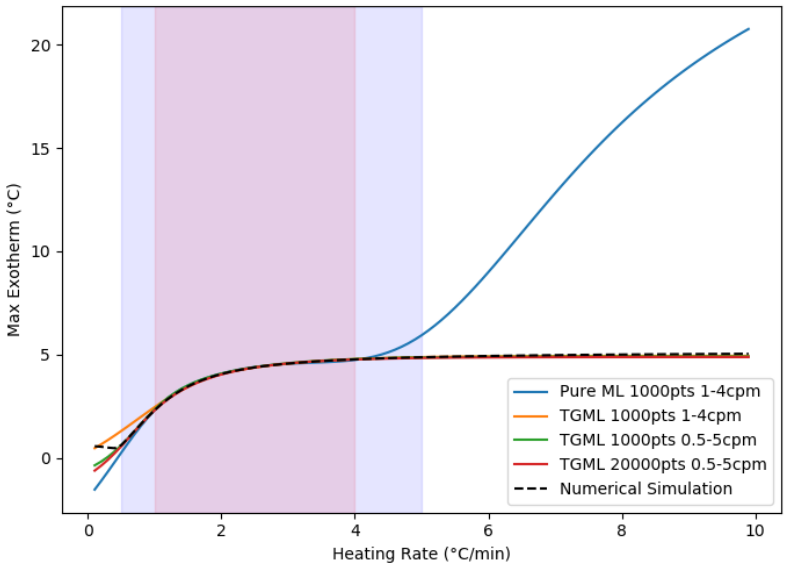

Figure 12: Unguided ("pure") ML vs TGML predictions for maximum exotherm for Invar tooling, L_2=10mm, h_1=60 W/m^{2}K, h_2=40 W/m^{2}K, L_1=8mm.

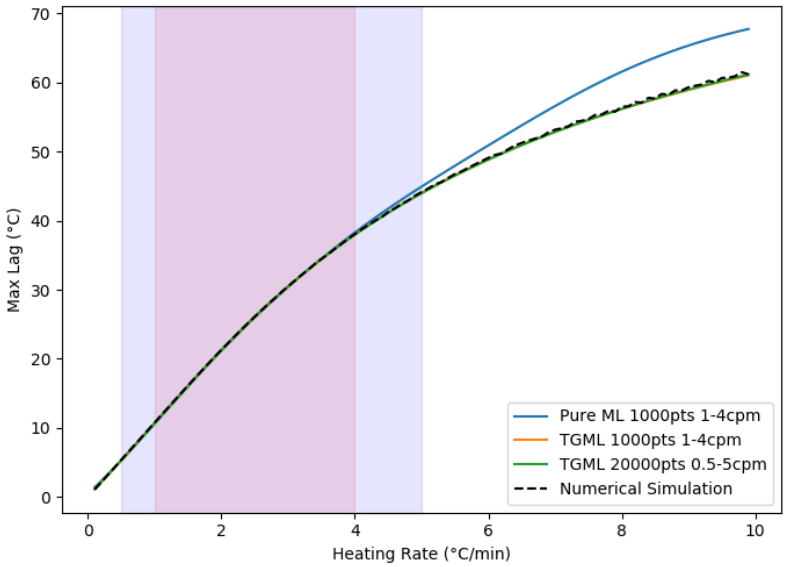

Figure 13: Unguided ("pure") ML vs TGML predictions for maximum lag for Invar tooling, L_2=10mm, h_1=60 W/m^{2}K, h_2=40 W/m^{2}K, L_1=8mm.

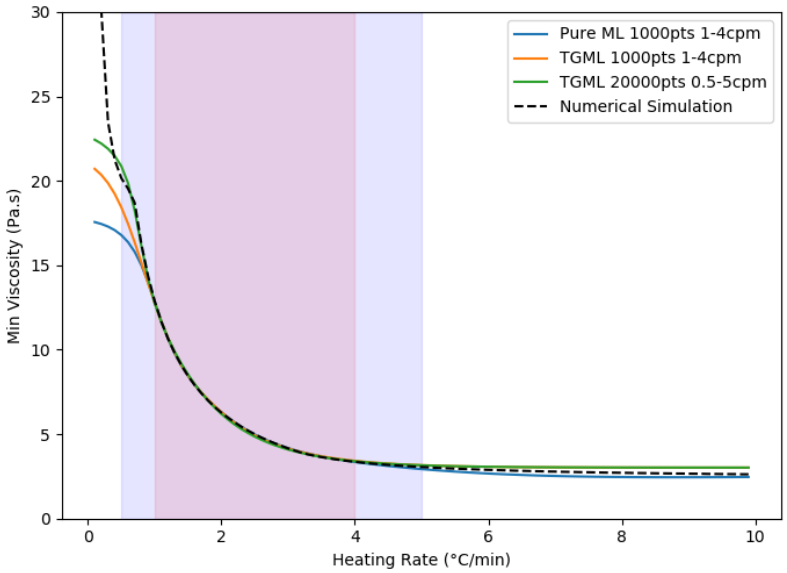

Figure 14: Unguided (“pure”) ML vs TGML predictions for minimum viscosity for Invar tooling, L_2=10mm, h_1=60 W/m^{2}K, h_2=40 W/m^{2}K, L_1=8mm.

In all the cases, the unguided ML and TGML approaches do a very good job at predicting the behaviour in the training zone, even with as few as 1000 training points. The TGML models do a much better job predicting the behaviour outside the training zone, particularly in the prediction of the exotherm and thermal lag KPIs.

Surprisingly few data points are required for training both the pure ML and TGML models. The tables below show the relationship between the number of training points and the error measurements.

Table 6: Unguided machine learning errors.

Training points	RMSE (°C)	Max Error (°C)
250	0.35	2.91
500	0.16	1.33
1000	0.11	1.23
5000	0.06	1.13
10000	0.07	0.74
20000	0.07	0.55

Table 7: Theory-guided machine learning errors.

Training points	RMSE (°C)	Max Error (°C)
250	0.16	1.07
500	0.11	1.0
1000	0.09	0.88
5000	0.06	0.50
10000	0.07	0.62
20000	0.06	0.54

Within the training zone, which is the only data range that the error measurements are applicable, the two approaches have results that are similar. TGML typically exhibits less maximum error for a given number of training points and somewhat better RMSE. For an acceptable maximum error of 1°C, as few as 500 training points are required for a TGML model, whereas a similar pure ML model requires 10,000 training points.

Furthermore, there is very little difference between training the TGML model with 5000 data points and 20,000 data points. Even using as few as 500 data points only changes the maximum error to 1°C from 0.5°C with 20,000 data points.

5. Use of TGML in Manufacturability Assessment of Preliminary Design

The TGML models discussed above were implemented in the CPA-TA software. These models served as a surrogate engine for the software, replacing the direct numerical analysis previously required to perform the manufacturability assessment. The TGML models generated the same results as the numerical simulations, but required only 20 seconds to complete. The runtimes of each of the analysis techniques is summarized in the table below.

Table 8: Runtime for each simulation technique.

Technique	Runtime
Full simulation	1620 s
+"Rounded" input parameters	276 s
+fully-cached results	38 s
TGML (no rounding or cache)	20 s

6. Summary

Use of theory-guided machine learning as a surrogate engine in composites process simulation has been demonstrated to be very effective in speeding up the time required to assess a preliminary design in a typical composites manufacturing case study. With the competing requirements of a fast design cycle, integration of physics-based simulation to reduce the number of pre-production prototypes, and increasing complexity in the design and manufacture of modern parts, simulation solutions must be found that are capable of returning accurate results in a timely manner. It has been shown that with an appropriate injection of processing science theory into the machine learning algorithms, a small number of training data can be used to predict a wide-range of response to a high degree of accuracy in a very short time. It is through this combination of intimate knowledge of the domain (composites manufacturing and process modelling in this case), and an understanding of the science of machine learning that efficient, physics-based solutions will be created for modern industrial design and manufacturing.

7. References

[1] Columbus, L. (2018). *Roundup of Machine Learning Forecasts and Market Estimates, 2018*, Retrieved 1 February 2019 from https://www.forbes.com/sites/louiscolumbus/2018/02/18/roundup-of-machine-learning-forecasts-and-market-estimates-2018

[2] Wagner, N., Rondinelli, J.M., (2016). *Theory-guided machine learning in materials science*, Frontiers in Materials, 3, 28.

[3] Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V. (2017). *Theory-guided data science: A new paradigm for scientific discovery from data.* IEEE Transactions on Knowledge and Data Engineering, 29(10), pp. 2318-2331.

[4] Buduma, N., Lacascio, N. (2017). *Fundamentals of Deep Learning.* O’Reilly Media, Inc.

[5] Convergent Manufacturing Technologies, Inc (2018), *RAVEN v3.9.7* [Computer software]

[6] Hubert, P., Johnston, A., Poursartip, A., Nelson, K. (2001). *Cure kinetics and viscosity models for Hexcel 8552 epoxy resin*, SAMPE Conference, Society for the Advancement of Material and Process Engineering, Seattle, WA, USA

[7] Zobeiry N., Poursartip A. (2019). *Theory-guided machine learning for processing of advanced composites*, Composites Part A: Applied Science and Manufacturing, submitted January 2019

[8] Zobeiry N., Park J., Poursartip A. (2018). *An IR Thermography based method for the evaluation of the thermal response of tooling for composites manufacturing*, Journal of Composite Materials, doi.org/10.1177/0021998318798444

[9] Rasekh, A., Vaziri, R., and Poursartip, A. (2004). *Simple Techniques for Thermal Analysis of the Processing of Composite Structures*. 36th International SAMPE Tech Conference, San Diego, CA, USA.

[10] Rasekh, A. (2007). *Efficient Methods for Non-Linear Thermochemical Analysis of Composite Structures Undergoing Autoclave Processing*, PhD Thesis, The University of British Columbia, Canada.