



# Machine learning and simulation-based surrogate modeling for improved process chain operation

André Hürkamp<sup>1</sup> · Sebastian Gellrich<sup>1</sup> · Antal Dér<sup>1</sup> · Christoph Herrmann<sup>1</sup> · Klaus Dröder<sup>1</sup> · Sebastian Thiede<sup>2</sup>

Received: 20 January 2021 / Accepted: 12 April 2021 / Published online: 19 May 2021  
© The Author(s) 2021

## Abstract

In this contribution, a concept is presented that combines different simulation paradigms during the engineering phase. These methods are transferred into the operation phase by the use of data-based surrogates. As an virtual production scenario, the process combination of thermoforming continuous fiber-reinforced thermoplastic sheets and injection overmolding of thermoplastic polymers is investigated. Since this process is very sensitive regarding the temperature, the volatile transfer time is considered in a dynamic process chain control. Based on numerical analyses of the injection molding process, a surrogate model is developed. It enables a fast prediction of the product quality based on the temperature history. The physical model is transferred to an agent-based process chain simulation identifying lead time, bottle necks and quality rates taking into account the whole process chain. In the second step of surrogate modeling, a feasible soft sensor model is derived for quality control over the process chain during the operation stage. For this specific uses case, the production rejection can be reduced by 12% compared to conventional static approaches.

**Keywords** Production engineering · Machine learning · Digital twin · Surrogate modeling · Process chain simulation · Cyber physical production systems · Overmolded thermoplastic composites

## 1 Introduction

In today's product development, it is a matter of course to support the product and production engineering by virtual methods. Model-based computer simulations are able to analyze the material behavior of products and processes using numerical methods based on finite elements or finite difference schemes. The utilization of virtual and digital methods reduces experimental testing and prototyping during the product development. Furthermore, sophisticated simulation models allow a deeper understanding of processes and structural behavior and in particular, their interaction. If additionally total life cycle impacts need to be assessed, integrated computational methods are the only available and adequate tool to analyze different production,

use and environmental scenarios with respect to the underlying complexity [1]. The more reliable the used methods and input data are, the more meaningful the scenario assessment will be.

In contrary to the model-based approaches used for engineering purposes, mainly data-based machine learning (ML) algorithms have proven their suitability for industrial operation purposes. By continuously collecting and processing sensor, machine and process data parallel to the manufacturing, these methods provide an objective, data-driven process optimization and failure detection [2–4]. In this context, digital twins are even able to implement a data-based control of the production [5, 6]. Several approaches for (virtual) quality gates in manufacturing systems based on process and machine data can be found for single processes, but very sparse approaches for process chains [7]. If process chains are considered, the data-driven models widely draw on product data such as intermediate and final product properties. In [8], a digital twin of a composite part is designed by means of sensor data and an analytical model. Only for likely critical parts additional finite element analyses were carried out to improve the digital twin. For injection molding, usually in-mold sensors and machine data are used for an adequate quality prediction

✉ André Hürkamp  
a.huerkamp@tu-braunschweig.de

<sup>1</sup> Technische Universität Braunschweig, Institute of Machine Tools and Production Technology, Braunschweig, Germany

<sup>2</sup> Faculty of Engineering Technology, Department of Design, Production & Management, Chair of Manufacturing Systems, University of Twente, Enschede, Netherlands

[9, 10]. There exist also recent approaches that do not require additional sensors by using simulation data from flow simulations in combination with neural networks to predict the quality of injection-molded components [11, 12]. However, these approaches concentrate mainly on the single process and not on a complete process chain.

Therefore, the objective of this paper is to demonstrate a general concept of how engineering tools and data (i.e., the knowledge of the system behavior and physical relations) can be transferred from the planning and engineering phase into the operation phase to serve as virtual quality gates or an adaptive process chain control. The proposed model transfer and the interaction of the simulation models is shown in Fig. 1.

The starting point of the proposed scheme is given by a high fidelity simulation using the finite element method (FEM). FEM is commonly used for a model-based engineering of production processes. Due to the underlying physical modeling of material properties and process kinematics, it is able to compute valid part properties. In this context, the Integrated Computational Materials Engineering combines methods from computational materials science and multi-scale mechanics considering the microstructure of materials and the corresponding effects of processing [13, 14]. Especially, in the field of composite manufacturing, virtual process chains are important to predict structural properties [15]. Furthermore, detailed simulations allow the prediction of internal temperature distributions that can hardly be measured to ensure the required temperature conditions during manufacturing [16, 17]. Since these simulations are usually time consuming and expensive to evaluate [18], a real-time suitable surrogate model is needed for virtual quality control to stay computational feasible.

According to [19], surrogate modeling is a technique that uses the sampled data to build surrogate models that are able to predict the output of an expensive computer

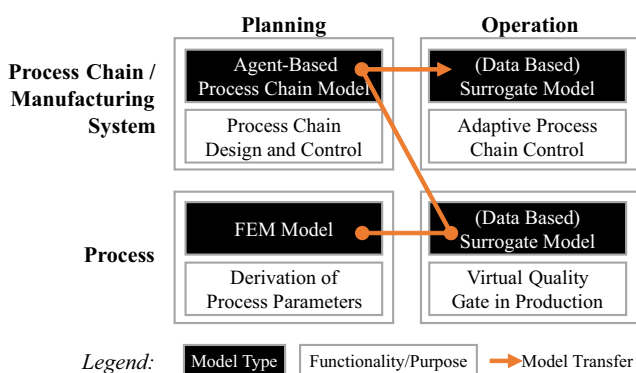
code at untried points in the parameter space. Within manufacturing, the data for surrogate modeling can be derived by structured parametric studies based on numerical simulations (process level) and/or agent-based/discrete event simulations (process chain/manufacturing system level). A general procedure for data-driven surrogate modeling of FEM data is proposed by Han and Zhang [19]. Applications can be found, e.g. in the field of biomechanics [20] or composite manufacturing [21]. An open issue in data-driven modeling of FEM data is its inevitable problem specificity, which is addressed by an approach of Zimmerling et al. [18]. In the context of this work, the surrogate predicts the resulting part properties based on the process parameters.

On higher abstraction levels, process chains and manufacturing systems are simulated by means of discrete-event (DE) and agent-based (AB) models [22]. With respect to the interdependencies between the single process steps of the process chain and interactions with further processes within a factory, the process chain modeling and simulation extends the scope of detailed process simulations. Extending the perspective from process to process chain and simultaneously integrating the insights from the process perspective allows for an evaluation of the whole process chain. In this way, the dynamic effects of changing a single process step (e.g. reduced cycle time or parallelization of machines) can be evaluated holistically. Furthermore, including surrogates from detailed process simulations in the process chain, simulation enables an integrated evaluation of the interactions of parameter variations [23]. Hence, it is possible to derive an optimal strategy for the control or evaluate the potential for the specific use case. By further deploying data-based surrogate models on the process chain model, an adaptive process chain control can be achieved.

The core idea of the concept is to derive an adaptive process chain control without extensive experimental data mining by using only engineering data. In the outline of this paper, the functionality of the concept is demonstrated on a purely virtual production scenario. However, it is also possible to combine the simulation-based approach with other data-driven approaches based on sensor and machine data in the operation.

## 2 Methods

The presented work is a continuation of preliminary works on Integrated Computational Product and Production Engineering (icPPE) [24] and surrogate modeling [17]. In the following, the main aspects will be discussed in terms of improved process chain operation.



**Fig. 1** Interaction between simulation models and their functionality in planning and operation

## 2.1 Integrated computational product and production engineering

Within the concept of icPPE, different simulation paradigms are combined along the multi-scale analysis of production systems focusing on both, the product engineering and its corresponding production processes. In Fig. 2, a schematic representation of icPPE is illustrated.

It distinguishes between the three levels product, process and process chain/ factory. At each level, different simulation methods are used with regard to the desired level of detail. Product properties such as mechanical performance (e.g. strength, stiffness) are mainly influenced by structural parameters (e.g. material orientation, wall thickness). These quantities are in turn mainly influenced by the manufacturing process. Therefore, the process parameters have a significant influence on the structural parameters and, thus, on the product properties. The relationship between process and structural parameters is investigated in a virtual process chain using, e.g. FEM. Since detailed FEM simulations are time-consuming at both, product and process level, surrogate (SG) models enable rapid parameter studies to be carried out with sufficient physical detail. In conjunction with machine data, the FEM data obtained at the process level is used as input for the process chain simulation. At the process chain/factory level, the physical detail of the processes is less relevant. Hence, mainly AB and DE simulations are used. On even higher aggregation levels, e.g. when a factory represents the lowest system level, also system dynamics (SD) is applied. Furthermore, the process chains impose requirements on the factory and vice versa.

As process and product quality are decisive for future competitiveness of manufacturing, icPPE strives to leverage data of engineering methods (FEM) and process chain

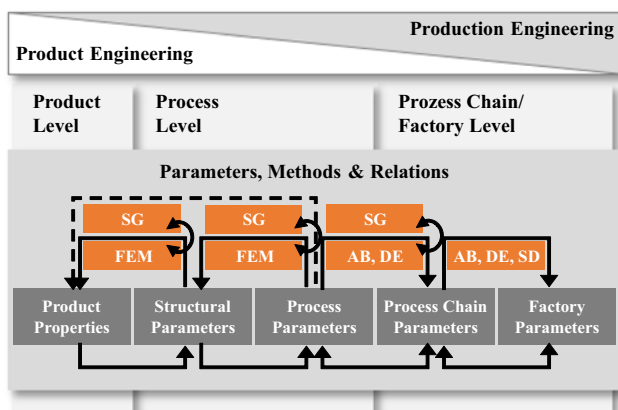
simulation, to improve manufacturing operation. This is accomplished by forwarding the detailed quality properties derived by FEM simulations along multiple scales of manufacturing. In this context, Hürkamp et al. show the applicability and potentials of fast and accurate data-driven surrogates of FEM simulations for the design of process chains [24] and as virtual quality gates [17]. In addition, Filz et al. [7] discuss the design and potentials of virtual quality gates for future manufacturing, considering FEM surrogate modeling as a promising approach.

## 2.2 Implementation concept for surrogate-based icPPE

The implementation of a surrogate-based icPPE for improved process chain operation is structured as two integrated cyber physical production systems (CPPS). In Fig. 3, the proposed concept is depicted. For demonstration purposes, it addresses the use case of overmolded thermo-plastic composite manufacturing described in Section 3. However, the implementation concept is general applicable transferable to other use cases.

The partition in two integrated CPPS is on the one hand due to the continuous interplay between physical world (process and process chain) and its cyber representative, i.e. surrogate model, and on the other hand, the integration of the quality-oriented process-specific surrogate model within the process chain simulation linking the two cycles. The complete concept covers eleven steps, which need to be done consecutively. Steps 1–9 focus on model training and engineering tasks and 10–12 on deployment of models respectively manufacturing operation. A detailed description on implementing a virtual quality gate on process level based on FEM data as shown in the left cycle in Fig. 3 (steps 1–4 and 10.1–12.1), can be found in [17]. In general, a structured design of experiment (DoE - 2) on a FEM simulation (1) is pursued, which serves as a training set for data-driven surrogate modeling. The surrogate modeling (3) is based on machine learning approaches, e.g. random forest and validated against experimental data. Finally, the surrogate is embedded within a decision support application, which is parameterized with sensor and machine data, as a virtual quality gate (10.1–12.1).

The process chain level expands the approach proposed in the previous study, by forwarding the surrogate model build on the process level to process chain simulations (5), e.g. integrated AB and DE simulation. Hereby, the approach exploits the surrogate models' property of fast and accurately predicting the product quality properties across the design parameter space. By defining thresholds for the amount or density of weak spots of the spatial surrogate model, a physics-based and process-parameter dependent quality rate can be derived, which is integrated



**Fig. 2** Concept of Integrated Computational Product and Production Engineering (icPPE) in accordance with [24]

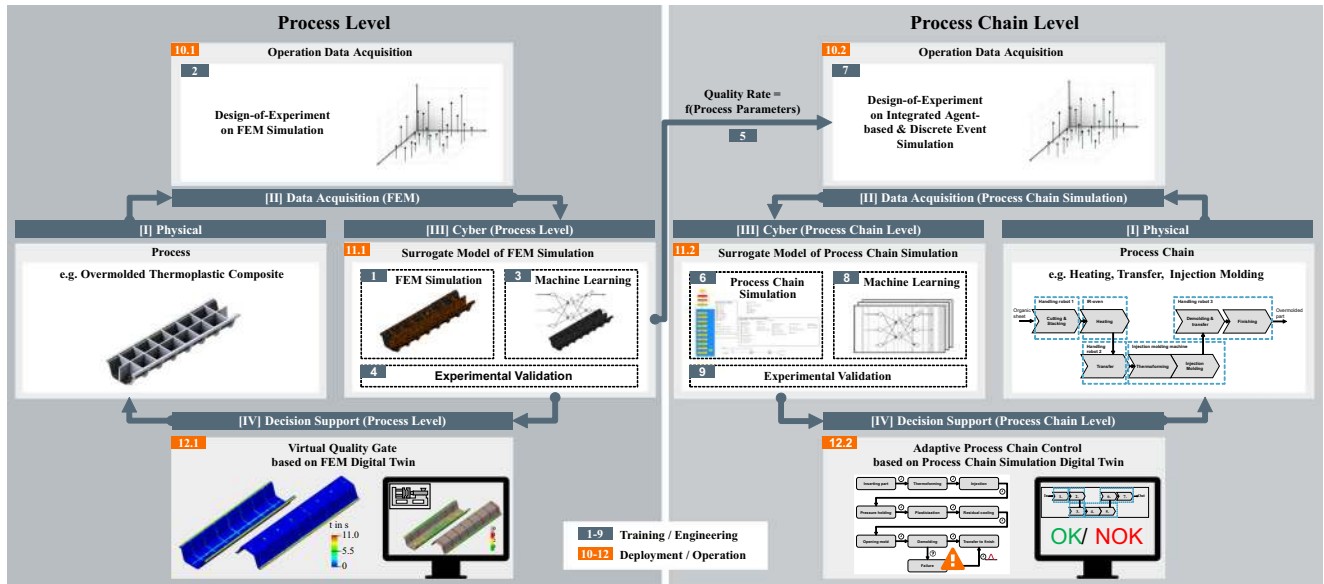


Fig. 3 Implementation concept for surrogate-based icPPE

as model parameter within the process chain simulation (5). As for the virtual quality gate on process level, a surrogate of a process chain simulation can be used as soft sensor that captures the dynamics of the underlying process chain, e.g. expected temperature of the part insert based on process chain dynamics. This information can be interpreted as decision support for adaptive process chain control measures (10.2–12.2), like reheating the organo sheet. In order to derive the soft sensor, a process chain simulation is carried out (6), taking into account the operational state specific quality rate modeled on process level. Analogous to process level surrogate modeling, a structured DoE on the process chain simulation (7) is calculated (e.g. change in ambient temperature, machine failures). Based on the simulation output, machine learning is applied with one or more target variables (8). Again, the soft sensor should be validated within an experimental setup or by supplementary measurements during production (9).

### 3 Use case study for the manufacturing of overmolded thermoplastic composites

Within this contribution, the expanded framework is demonstrated by means of a manufacturing system for the production of overmolded thermoplastic composites. Overmolded thermoplastic composites are produced in an integrated manufacturing process, combining thermoforming of continuously fiber reinforced thermoplastic composite sheets (organo sheets) and injection molding of a thermoplastic polymer that is compatible with the organo sheet’s matrix [25, 26]. In that way, composite structures with high specific

stiffness and strength can be produced on an injection molding machine, which allows an economic production with short cycle times.

The quality of the final structure is determined by the resulting bond strength between organo sheet and injected polymer since it determines to a large extend its ultimate strength. Experimental results emphasize that this interface bond strength depends mainly on the interface temperature during the process, which is also considered in different modeling approaches [17, 27, 28]. According to [17], the contact time

$$t_c = \int T_{if} \geq T_{PP}^* dt, \tag{1}$$

is a suitable process indicator for the final part quality. It is defined by the time the interface temperature  $T_{if}$  is greater than the melting temperature  $T_{PP}^* = 163^\circ\text{C}$ . Since the contact time  $t_c$  is hardly to measure within the process, it is computed from injection molding simulations using the

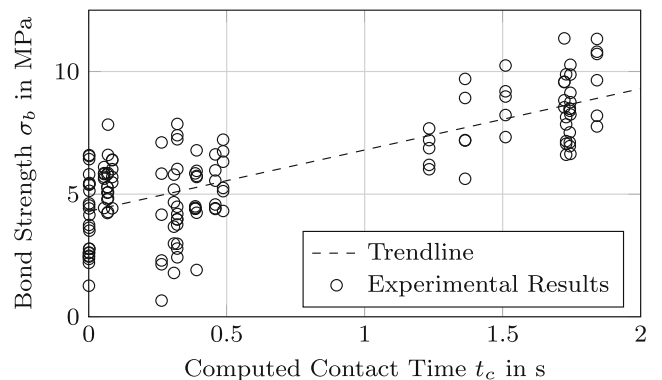
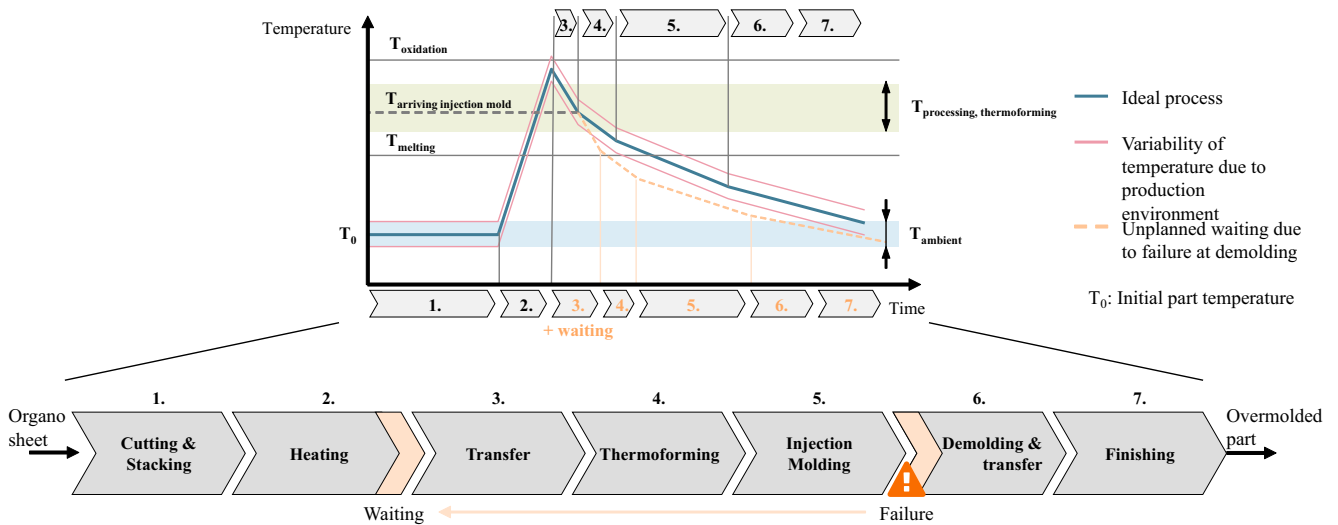


Fig. 4 Correlation between contact time  $t_c$  and bond strength  $\sigma_b$  [17]



**Fig. 5** Process chain and corresponding interface temperature of overmolded thermoplastic composite manufacturing

same process parameters as for the manufacturing of the testing specimens. In Fig. 4, the correlation between contact time and experimentally determined bond strength values of cross tension specimens is depicted. Fig. 5.

It can be seen that the bond strength is strongly correlated to the contact time and that larger contact times result in higher bond strength values. Hence, the part quality strongly correlates with the contact time.

### 3.1 Process chain for overmolded thermoplastic composites

A schematic illustration of the process chain for manufacturing overmolded thermoplastic composites and the corresponding interface temperature  $T_{if}$  is shown in starting from the cutting and stacking (1.), the organo sheets are heated up from the initial temperature ( $T_0$ ) to the processing temperature (2.), which is usually above the melting temperature  $T_{melting}$  of the matrix and below the oxidation temperature  $T_{oxidation}$ . The heating can either be carried out in an infrared oven or directly by an in-mold heating device. The heated organo sheet is then transferred (3.) by, e.g. a handling robot to the injection molding machine, where the sheet starts to cool down due to heat convection. Hereby, it has to be ensured that the organo sheet’s temperature when leaving the oven is high enough to compensate the cooling [29]. After reaching the final position, the organo sheet is thermoformed by closing the mold (4.) and directly afterwards overmolded in the same tool (5.). After a cooling phase and solidification in the mold, the part is demolded (6.) and transferred to finishing (7.). In addition to the ideal temperature distribution (blue curve), a variability in the course of the temperature might occur due to different temperature influences in the production

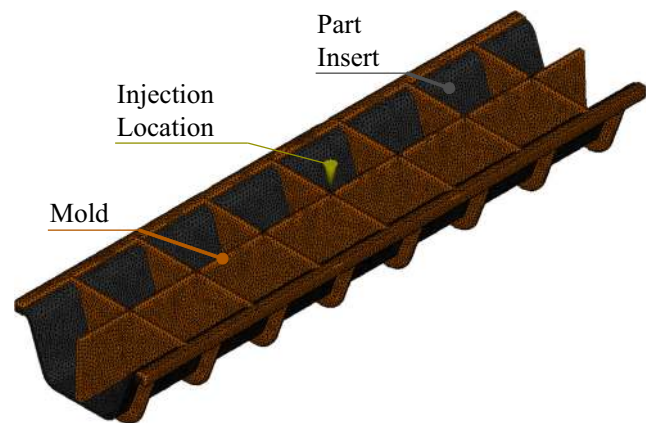
zone. Furthermore, possible failures along the process chain lead to waiting times that cause too much cooling to ensure sufficient product qualities.

### 3.2 Set up of the parametric study

The proposed concept is exemplary investigated for the structure displayed in Fig. 6.

For the numerical studies, the cavity and the part insert are discretized by 872,443 4-node tetrahedral elements. The injection location is placed at the center of the structure. The numerical parametric study of the injection molding is performed in Autodesk Moldflow using polypropylene as injection polymer. All necessary material data are taken from the Autodesk Moldflow material database [30].

The sampling space is designed in accordance with [17]. In order to achieve a meaningful database of different



**Fig. 6** Tetrahedral mesh of the overmolded thermoplastic structure used in the numerical studies



process combinations with minimal computational effort, a latin hypercube sampling [31] with  $n = 100$  samples is used. The minimum and maximum values and the corresponding distribution of the process parameters used in the parametric study are summarized in Table 1.

The melt temperature of the injected polymer  $T_{inj} = 240$  °C is constant for all simulations. For each of the 100 full-scale simulations, the contact time  $t_c$  (1) is computed from the temperature evolution during the simulation and stored in a database for the subsequent surrogate modeling.

### 3.3 Data-driven surrogate modeling of FE-simulation

In order to run a comprehensive study on quality properties across the possible design space of process parameters, a data-driven surrogate model of FE simulation is build. The surrogate model is trained based on the sampling data set derived in Section 3.2, taking 80 simulations as training data and 20 simulations as testing data. The final surrogate model is then deployed on 1000 further samples across the parameter space to provide a sophisticated data set for quality analysis. Due to the importance of the interface bond strength on the final product quality, as output feature of the surrogate model the spatial distribution of  $t_c$  is considered. As input features, the surface of the part insert (i.e. tetrahedral mesh with nodal coordinates  $x, y, z$  encompassing 25,351 nodes), the flow rate, mold temperature and part insert temperature are chosen. In order to stay computational feasible, the training data is down-sampled to 20%, which results in a final training data set that is sized  $405,600 \times 7$ . In contrast, no down-sampling is applied for the testing data, having a test set dimension of  $507,020 \times 7$ . Within data pre-processing, the features are standardized by MinMax scaling in the range of 0 to 1. Within previous studies, the two data-driven approaches decision tree (DT), i.e. good fit and fast training / prediction, and random forest (RF), i.e. best fit, have shown best performances for surrogate modeling of FE simulation data [17, 24, 32]. Hence, within this paper, DT and RF are selected for the data-driven modeling. The modeling is implemented based on the Python library scikit-learn [33] using a 5-fold cross-validation. The model parameters of

**Table 1** Investigated process parameters and the corresponding sampling distribution

Process parameter	Min	Max	Distribution
Part insert Temperature in °C	20	240	Modified Log-normal
Mold temperature in °C	30	80	Uniform
Flow rate in $cm^3/s$	10	100	Uniform

**Table 2** Hyperparameter optimization for DT and RF

Method	Model parameter	Range	# Steps	Optimum
Decision tree	Max depth	1–none	7	200
	Min samples leaf	1–10	10	3
Random forest	Max depth	10–110	6	50
	Min samples leaf	1–5	3	1
	# estimators	10–900	8	900

both methods are tuned according to the parameters listed in Table 2. Here, a wider parameter spectrum is investigated than in previous studies.

The best-found model parameters in parameter tuning for  $R^2$  are shown in the last column of Table 2. Most of the parameters (except of amount of estimators for RF) lie within the defined parameter range, so that an under- and overfitting should have not been occurred. This is also reflected within the test data metrics shown in Table 3. The table shows the scores based on the test data set including their standard deviation. The metrics  $R^2$ , mean absolute error (MAE), mean squared error (MSE) and mean maximum error (Mean Max Error) reveal the slightly better performance and in general smaller standard deviation of RF against DT. However, DT is much faster in training and prediction than RF, which becomes crucial for real-time applicability as virtual quality gate in production, especially for complex parts. For both models, the smallest maximum error on test data is achieved on sample 96 manufactured under the process conditions of part insert temperature = 158.85 °C, mold temperature = 48.25 °C and flow rate = 76.15  $cm^3/s$ . Again, RF reaches a better performance with 2.1248 s as minimum max error in contrast to 4.4306 s for DT. However, such larger errors appear very rarely, which should not affect the overall decision on product quality.

### 3.4 Quality analysis

The simulation and consequently the surrogate model yields a detailed distribution of the contact time  $t_c$ . Here, we define the quality of a part based on the quality domains defined by

**Table 3** Test data metrics for FEM surrogate modeling based on DT and RF (test score  $\pm$  standard deviation)

Metric	Decision tree	Random forest
$R^2$	0.9868 $\pm$ 0.0086	0.9919 $\pm$ 0.0077
MAE	0.1278 $\pm$ 0.0539	0.1028 $\pm$ 0.0549
MSE	0.2199 $\pm$ 0.1689	0.1338 $\pm$ 0.1425
Mean Max Error	7.3249 $\pm$ 1.8929	5.4323 $\pm$ 1.7371
Training time	1.39 s	133.37 s
Prediction time	0.05 s	9.27 s

**Table 4** Classification of quality values

Numerical value	Range	Classification
$q = 0$	–	“no bond”
$q = 1$	$t_c \leq 0.4$	“poor”
$q = 2$	$0.4 < t_c \leq 1.5$	“good”
$q = 3$	$t_c > 1.5$	“excellent”

Hürkamp et al. [17]. For each value of  $t_c$ , a discrete numeric value  $q$  is defined in order to distinguish between “poor”, “good” and “excellent” bond strength values. The discrete values for the present use case are given in Table 4. The nodes of the part insert without interface to the polymer are denoted by  $q = 0$ .

In order to derive valuable quality criteria for the whole part, a specific quality measure needs to be defined that evaluates if a part will be “okay” (OK) or “not okay” (NOK). Here, we define the part quality

$$Q_{\text{part}} = 1 - \frac{|\{n|q(n) = 1\}|}{n_{\text{tot}}} \quad \text{with} \quad 0 \leq Q_{\text{part}} \leq 1 \quad (2)$$

by the number of nodes that have a poor ( $q=1$ ) quality. A value of  $Q_{\text{part}} = 1$  represents a part of perfect quality. If the areas with poor quality are too large, the structural integrity is low and hence, the total part quality will be low.

In Fig. 7, five exemplary quality distributions are illustrated. It can be seen that the number of poor values decreases with increasing temperature. Furthermore, it is observed that the quality inside the U-profile is always better than on the outside. This is due to the fact that the flow path increases and therefore the part insert and the plastic melt exhibit lower temperatures and thus, lower bond strength.

### 3.5 Process chain simulation

In order to address the challenges associated with an integrated process-process chain modeling and to provide the needed insights, an integrated AB and DE modeling and simulation approach is chosen. Core of the modeling approach is the material flow in the sense of a passive DE process chain model. The AB modeling principles are consequently applied to extend the passive material flow by active objects and their behavior, i.e. machines and products. Going beyond the logic of passive product entities that only enter and exit process steps in a conventional DE model, product agents store a set of parameter values that interact bidirectionally with machine models. Based on the parameter values stored in a product agent, modeling an adaptive machine control becomes feasible (e.g. product-specific heating time based on arrival temperature before an oven). Machines change in turn during their processing

sequence the product agent’s parameter values (e.g. part temperature) or store additional information in the agent (e.g. part quality).

Within this case study, the process chain model covers the process steps from Fig. 5. Note that the present use case represents a rather linear process chain for demonstrations purposes. The subject to be assessed was how process chain inherent dynamics affect the quality rate of the overmolding process. The most critical factor for the quality of the process is the surface temperature of the part insert at the injection molding machine. The heating cycle inside the infrared oven is modeled with a simplified physics-based approach, where the end temperature of the organo sheet

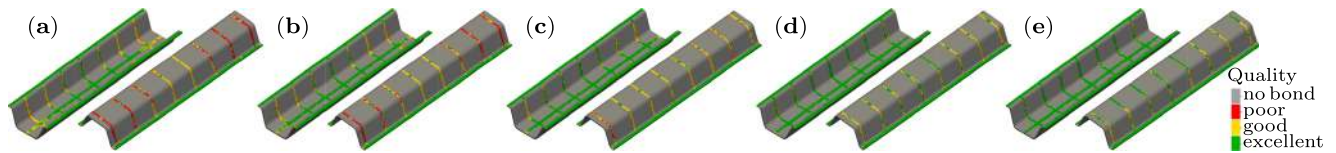
$$T_1(t) = \frac{P \cdot t}{m \cdot c} + T_0 \quad (3)$$

is calculated via its mass  $m$ , its specific heat capacity  $c$ , the initial temperature  $T_0$  and a constantly applied heating power  $P$ . The oven is controlled according to the ideal overmolding cycle without any failures. In this case, the heated organo sheet is directly transferred into the mold. During this time (transfer and insertion into the mold), the organo sheet starts to cool down. Its cooling behavior is approximated based on an empirical cooling curve that was measured and parametrized. The current temperature  $T(t)$

$$T(t) = (T_1 - T_A) \cdot e^{-at} + T_A \quad (4)$$

depends on the cooling rate  $a = 0.011 \text{ s}^{-1}$ , the ambient temperature  $T_A$  and the part temperature after the heating cycle  $T_1$ . At this point of the process chain simulation, the FEM-surrogate model calculates the part quality based on the arrival temperature at the mold.

For the use case, ten scenarios, each with 10,000 parts were designed. Within this factorial scenario assessment five different ambient temperatures of 10, 15, 20, 25 and 30 °C in the production zone and two different machine failure probabilities of 5% and 10% are combined. The ambient temperature in the production zone can vary seasonally (e.g. in summer and winter) and more dynamically over a day (e.g. open gates upon a delivery). These fluctuations are also accompanied by the cooling process during the transfer from the oven into the injection mold. Although, injection molding is an established manufacturing process for high production volumes, the process is subjected to infrequent failures. For the sake of simplicity, it is assumed that failures only occur during demolding. Those failures in the injection molding machine model are attributed to clamping or sticking to the wrong side of the tool. The resulting time delay at the injection molding machine causes waiting times at the oven after the heating cycle, which consequently leads to a longer cooling time and lower arrival temperature at the mold. Additionally, the part temperature before heating varies, which in turn affects the temperature after the



**Fig. 7** Exemplary illustration of different qualities based on the computed contact time  $t_c$  and corresponding quality values computed from the RF surrogate model: **a**  $T_{in} = 180.28^\circ\text{C}$ ,  $Q_i = 0.8807$ ;

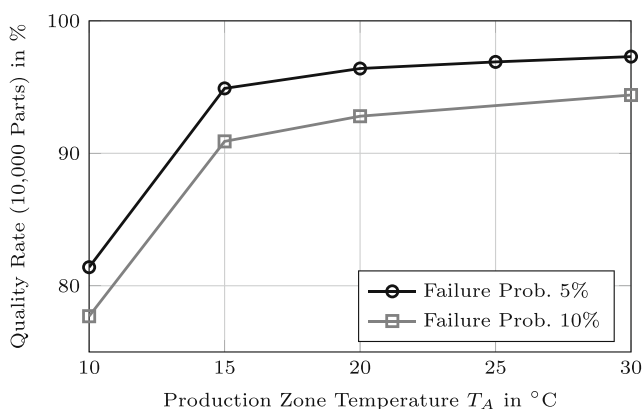
**b**  $T_{in} = 186, 03^\circ\text{C}$ ;  $Q_i = 0.9062$ , **c**  $T_{in} = 188.80^\circ\text{C}$ ,  $Q_i = 0.9636$ ;  
**d**  $T_{in} = 205.53^\circ\text{C}$ ,  $Q_i = 0.9895$ ; **e**  $T_{in} = 219, 09^\circ\text{C}$   $Q_i = 0.9929$

heating (with constant  $P$  and  $t$ ) cycle. These fluctuations are for example attributed to direct sunlight exposure, closeness to a heat source (e.g. the oven itself) or storage conditions.

In Fig. 8, the quality rate is determined for each scenario with respect to the ambient temperature and the failure probability. Here, we observe a significant increase of the quality rate when the mean temperature in the production zone increases from 10 to  $15^\circ\text{C}$ . This behavior indicates the temperature sensitivity of the production process. Based on the temperature of the organo sheet when arriving at the oven and the corresponding waiting times, the product quality is predicted. In particular, it is determined if a part will be OK or NOK. In Fig. 9, a map of the resulting qualities for the standard scenario with  $20^\circ\text{C}$  mean ambient temperature and a failure rate of 5% is depicted.

From the results, it can be seen that for a perfect process (total time between leaving oven and starting the thermoforming process is 12 s) with a defined heating power of the oven, the product's quality is ensured almost temperature independent. Only for relatively cold sheets below  $3.5^\circ\text{C}$ , the quality criterion is not fulfilled. Furthermore, an increase in tolerable transfer times is observed with increasing temperatures.

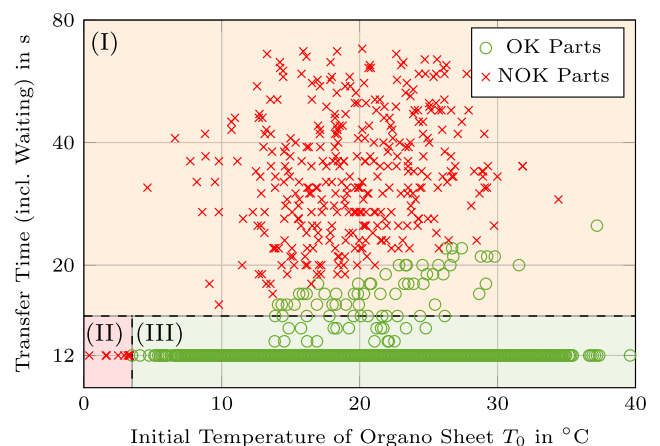
Without the scenario assessment, a constant threshold would probably be introduced to sort out parts after a significant downtime. In Fig. 9, an exemplary fixed threshold of 15 s is drawn to show the potential of the proposed dynamic, physical modeling. By this threshold,



**Fig. 8** Quality rate vs. production zone temperature for failure probabilities of 5% and 10%

the quality map is divided into three areas. All parts with a larger transfer time than the threshold are contained in area (I). Without the physical knowledge from the surrogate, all parts in this area would be denoted as failure parts and they would have been sorted out before further processing. Since the surrogate predicts the quality based on the complete part temperature history, also larger waiting times would lead to OK parts, when the initial temperature was large enough. In this particular example, 12% of the discarded parts would still be OK. In addition, we observe in area (II) only NOK parts, although they are under the threshold. Hence, a fixed threshold could also lead to NOK parts without rejecting defective parts before processing. Finally, only the parts contained in area (III) would be denoted as OK parts when a fixed threshold for the transfer time is used as control variable. In comparison with the total number of OK parts, this would extremely narrow down the process window. Although, the present use case study represents a rather linear process chain, the installation of physical information shows a significant impact on the number of accepted or rejected parts.

The case study underlines the potential of using the dynamic process chain surrogate modeling approach: neglecting the (potentially volatile) transfer time as important state variable leads to a significant share of NOK parts (3.6% in the investigated use case). Although, using a fixed threshold for the transfer can help to reduce the share of



**Fig. 9** Quality analysis for a virtual production scenario with mean  $T_0 = 20^\circ\text{C}$  and failure probability of 5%



NOK parts, it leads to a significant amount of parts that are falsely declared as NOK which eventually results in an inefficient utilization of resources.

### 3.6 Transfer into operation

In order to enable an adaptive process chain control (cf., Fig. 1 top right), a model-based soft sensor for process chain control is proposed by transferring the process chain simulation into operation. This is done analogously to the surrogate based transfer of FEM simulation into operation as a virtual quality gate by data-driven surrogate modeling (cf. Fig. 3, steps 8–12). Drawing on the process chain DoE in the previous subsection, a solid data base for surrogate modeling is compiled. The data set covers the above-described 10 scenarios. Besides the ambient temperature and failure probability, the temperature at leaving the oven as well as the transfer time is considered as input features. Those features are easy to measure in a real production environment, which makes them appropriate for the parametrization of the soft sensor to predict the expected product quality before actual processing of the organo sheet. For demonstrative purposes, a simple model benchmark without intensive feature tuning is carried out. The benchmark encompasses the models DT (max depth: None; min samples per leaf: 2), logistic regression (LR; epochs: 400; epsilon:  $1E-5$ ; learning rate: line search) and RF (estimators: 100; no restrictions on tree depth and minimum samples per leaf). For model evaluation a train-test-split, i.e.  $.8/.2$ , stratified on product quality due to high class imbalance is done. The final test set includes 57 NOK parts within the summer scenario ( $30^{\circ}\text{C}$  ambient temperature, 5% failure probability) and 437 NOK parts at winter scenario ( $10^{\circ}\text{C}$  ambient temperature, 10% failure probability).

The surrogate modeling results listed in Table 5 indicate a high classification accuracy, with no misclassifications for DT and RF on test data and only a few false positives for the OK class by the LR approach. According to the LR model, those 32 misclassified parts would have been

manufactured, although it is likely to result in a bad product quality. Those surrogate models, which still have to be experimentally validated, can be deployed as a soft sensor for active decision support on process chain control based on only a few easy to measure and/or calculate variables.

## 4 Conclusion

Within this paper, a novel concept of combining and transferring different engineering and planning simulation paradigms into the operation stage by data-driven surrogate modeling is presented for the production of overmolded thermoplastic composites. Based on a detailed parametric study of injection molding simulation, a data-driven surrogate is developed that is able to predict physical reliable product qualities based on the given process parameters. Hereby, RF performs slightly better than DT. However, DT is much faster in training and prediction than RF. In order to assess different production scenarios, the surrogate is transferred into a process chain simulation to derive process parameter-dependent quality rates. From the results, it is observed, that for cold production zones (below  $15^{\circ}\text{C}$ ) the quality rate significantly decreases. This is an interesting finding to be considered for the production in regions with strong seasonal variations in temperature. In addition to the planning support, the model is again transferred into the operation stage by a second step of surrogate modeling. Once more a parametric study is conducted, this time based on process chain simulation. Based on the parametric study, a soft sensor for adaptive process chain control is derived. The soft sensor captures process chain dynamics like machine failures and transfer times. In the investigated use case example the dynamic model was able to reduce the production reject of 12%. Furthermore, the dynamic model enables a real-time decision on material processing, e.g. reheating, by model-based prediction of the expected product quality based on easy to measure variables. Hence, it leads to a better utilization of production resources. The studied models (DT, RF, LogR) show an overall good performance.

In the present use case, the part quality is limited to the evaluation of only one variable (contact time  $t_c$ ). In real-world production environments, also other variables such as pressure influence the quality and need to be introduced in future works. Furthermore, an improvement of the surrogate modeling may be achieved by adding model reduction techniques for the data generation and pre-processing. Eventually, the developed concept needs to be tested and experimentally validated.

**Table 5** Metrics for test data of process chain surrogate modeling

Metric	DT/RF		LogR	
	OK	NOK	OK	NOK
True pos.	18321	1679	18321	1647
False pos.	0	0	32	0
Precision	1.000	1.000	0.998	1.000
Recall	1.000	1.000	1.000	0.981
Accuracy	1.000		0.998	

In summary, this contribution shows the potential of transferring already well-established methods for engineering purposes into the operation stage. With sophisticated simulation models, the main features of the system behavior can be implemented into CPPS already during product development without large experimental or sensorization effort. After start of production, of course, it can then be further improved by continuously accruing production data.

**Author contribution** A. Hürkamp, conceptualization, methodology, writing—original draft preparation, review and editing, visualization; S. Gellrich, conceptualization, methodology, writing—original draft preparation, visualization; A. Dér, methodology, writing—original draft preparation, visualization; C. Herrmann, resources, writing—review and editing, supervision, funding acquisition; K. Dröder, resources, writing—review and editing, supervision, funding acquisition; S. Thiede, conceptualization, methodology, writing—original draft preparation, review and editing, visualization.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This research and the results published are based on the research program Mobilise funded by the Ministry of Science and Culture of Lower Saxony and the Volkswagen Foundation. The support and sponsorship is gratefully acknowledged and appreciated.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Cerdas F, Thiede S, Herrmann C (2018) Integrated computational life cycle engineering — application to the case of electric vehicles. *CIRP Ann* 67(1):25–28. <https://doi.org/10.1016/j.cirp.2018.04.052>
- Wuest T, Irgens C, Thoben KD (2014) An approach to monitoring quality in manufacturing using supervised machine learning on product state data. *J Intell Manuf* 25(5):1167–1180. <https://doi.org/10.1007/s10845-013-0761-y>
- Ademujimi TT, Brundage MP, Prabhu VV (2017) A review of current machine learning techniques used in manufacturing diagnosis. In: Lödding H, Riedel R, Thoben KD, von Cieminski G, Kiritsis D (eds) *Advances in production management systems. The path to intelligent, collaborative and sustainable manufacturing*. Springer International Publishing, Cham, pp 407–415
- Weichert D, Link P, Stoll A, Rüping S, Ihlenfeldt S, Wrobel S (2019) A review of machine learning for the optimization of production processes. *Int J Adv Manuf Technol* 104(5):1889–1902. <https://doi.org/10.1007/s00170-019-03988-5>
- Gao RX, Wang L, Helu M, Teti R (2020) Big data analytics for smart factories of the future. *CIRP Ann* 69(2):668–692. <https://doi.org/10.1016/j.cirp.2020.05.002>
- Stark R, Damerou T (2019). In: Chatti S, Tolio T (eds) *CIRP encyclopedia of production engineering*. Springer, Berlin, pp 1–8. [https://doi.org/10.1007/978-3-642-35950-7\\_16870-1](https://doi.org/10.1007/978-3-642-35950-7_16870-1)
- Filz MA, Gellrich S, Turetskyy A, Wessel J, Herrmann C, Thiede S (2020) Virtual quality gates in manufacturing systems: framework, implementation and potential. *J Manuf Materials Process* 4. <https://doi.org/10.3390/jmmp4040106>
- Zambal S, Eitzinger C, Clarke M, Klintworth J, Mechin P (2018) A digital twin for composite parts manufacturing : effects of defects analysis based on manufacturing data. In: 2018 IEEE 16th International conference on industrial informatics (INDIN), pp 803–808. <https://doi.org/10.1109/INDIN.2018.8472014>
- Chen JC, Guo G, Wang WN (2020) Artificial neural network-based online defect detection system with in-mold temperature and pressure sensors for high precision injection molding. *Int J Adv Manuf Technol* 110(7):2023–2033. <https://doi.org/10.1007/s00170-020-06011-4>
- Farahani S, Brown N, Loftis J, Krick C, Pichl F, Vaculik R, Pilla S (2019) Evaluation of in-mold sensors and machine data towards enhancing product quality and process monitoring via Industry 4.0. *Int J Adv Manuf Technol* 105(1):1371–1389. <https://doi.org/10.1007/s00170-019-04323-8>
- Song Z, Liu S, Wang X, Hu Z (2020) Optimization and prediction of volume shrinkage and warpage of injection-molded thin-walled parts based on neural network. *Int J Adv Manuf Technol* 109(3):755–769. <https://doi.org/10.1007/s00170-020-05558-6>
- Lockner Y, Hopmann C (2021) Induced network-based transfer learning in injection molding for process modelling and optimization with artificial neural networks. *Int J Adv Manuf Technol* 112(11):3501–3513. <https://doi.org/10.1007/s00170-020-06511-3>
- Allison J, Backman D, Christodoulou L (2006) Integrated computational materials engineering: a new paradigm for the global materials profession. *JOM* 58(11):25–27. <https://doi.org/10.1007/s11837-006-0223-5>
- Panchal JH, Kalidindi SR, McDowell DL (2013) Key computational modeling issues in integrated computational materials engineering. *Comput Aided Des* 45(1):4–25. <https://doi.org/10.1016/j.cad.2012.06.006>
- Kärger L, Bernath A, Fritz F, Galkin S, Magagnato D, Oeckerath A, Schön A, Henning F (2015) Development and validation of a CAE chain for unidirectional fibre reinforced composite components. *Compos Struct* 132:350–358. <https://doi.org/10.1016/j.compstruct.2015.05.047>
- Kazan H, Farahani S, Pilla S (2019) Feasibility study for manufacturing CF/Epoxy – thermoplastic hybrid structures in a single operation. *Procedia Manufacturing* 33:232–239. <https://doi.org/10.1016/j.promfg.2019.04.028>
- Hürkamp A, Gellrich S, Ossowski T, Beuscher J, Thiede S, Herrmann C, Dröder K (2020) Combining simulation and machine learning as digital twin for the manufacturing of overmolded thermoplastic composites. *J Manuf Mater Process* 4(3):92. <https://doi.org/10.3390/jmmp4030092>
- Zimmerling C, Poppe C, Kärger L (2020) Estimating optimum process parameters in textile draping of variable part geometries - a reinforcement learning approach. *Procedia Manufacturing* 47:847–854. <https://doi.org/10.1016/j.promfg.2020.04.263>
- Han ZH, Zhang KS (2012) Surrogate-based optimization. In: Roeva O (ed) *Real-world applications of genetic algorithms*, InTech

20. Liang L, Liu M, Martin C, Sun W (2018) A deep learning approach to estimate stress distribution: a fast and accurate surrogate of finite-element analysis. *J R Soc Interface* 15(138). <https://doi.org/10.1098/rsif.2017.0844>
21. Pfrommer J, Zimmerling C, Liu J, Kärger L, Henning F, Beyerer J (2018) Optimisation of manufacturing process parameters using deep neural networks as surrogate models. *Procedia CIRP* 72:426–431. <https://doi.org/10.1016/j.procir.2018.03.046>
22. Borshchev A, Filippov A (2004) From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. In: *Proceedings of the 22nd international conference of the system dynamics society*, vol 22, pp 25–29
23. Herrmann C, Heinemann T, Thiede S (2011) Synergies from process and energy oriented process chain simulation – a case study from the aluminium die casting industry. In: Hesselbach J, Herrmann C (eds) *Glocalized solutions for sustainability in manufacturing*. Springer, Berlin, pp 317–322
24. Hürkamp A, Dé A, Gellrich S, Ossowski T, Lorenz R, Behrens BA, Herrmann C, Dröder K, Thiede S (2020) Integrated computational product and production engineering for multi-material lightweight structures. *Int J Adv Manuf Technol* 110(9–10):2551–2571. <https://doi.org/10.1007/s00170-020-05895-6>
25. Bouwman M, Donderwinkel T, Krämer E, Wijskamp S, Costa SF (2016) Overmoulding—an integrated design approach for dimensional accuracy and strength of structural parts. *ITHEC Proceedings*
26. Valverde MA, Kupfer R, Wollmann T, Kawashita LF, Gude M, Hallett SR (2020) Influence of component design on features and properties in thermoplastic overmoulded composites. *Compos A: Appl Sci Manuf* 132:105823. <https://doi.org/10.1016/j.compositesa.2020.105823>
27. Akkerman R, Bouwman M, Wijskamp S (2020) Analysis of the thermoplastic composite overmolding process: interface strength. *Frontiers in Materials* 7. <https://doi.org/10.3389/fmats.2020.00027>
28. Giusti R, Lucchetta G (2020) Modeling the adhesion bonding strength in injection overmolding of polypropylene parts. *Polymers* 12. <https://doi.org/10.3390/polym12092063>
29. Bruns C, Tielking JC, Kuolt H, Raatz A (2018) Modelling and evaluating the heat transfer of molten thermoplastic fabrics in automated handling processes. *Procedia CIRP* 76:79–84. <https://doi.org/10.1016/j.procir.2018.01.011>
30. Autodesk (2017) *Moldflow Insight*
31. McKay MD, Beckman RJ, Conover WJ (1979) Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21(2):239–245. <https://doi.org/10.1080/00401706.1979.10489755>
32. Swischuk R, Mainini L, Peherstorfer B, Willcox K (2019) Projection-based model reduction: formulations for physics-based machine learning. *Computers & Fluids* 179:704–717. <https://doi.org/10.1016/j.compfluid.2018.07.021>
33. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825–2830

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.